

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:14 ; Search time 15.0327 Seconds  
(without alignments)  
1470.857 Million cell updates/sec

Title: US-09-840-743-72

Perfect score: 693

Sequence: 1 WXXXXXXXXXXXXXXXXXX.....XXXXXXXXXXXXXXXXXXXX 230

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	228	32.9	555	2 T48453	hypothetical prote
2	201.5	29.1	917	2 T05430	hypothetical prote
3	139	20.1	1207	2 D84781	hypothetical prote
4	112.5	16.2	259	2 D75275	endonuclease III -
5	89.5	12.9	236	2 S75373	probable DNA-(apur
6	88.5	12.9	265	2 D84198	endonuclease III f
7	86.5	12.5	241	2 G87530	endonuclease III f
8	86	12.4	222	2 A75109	endonuclease III f
9	83.5	12.0	233	2 B69202	endonuclease III -
10	82	11.8	222	2 F71025	endonuclease III -
11	78.5	11.3	338	2 C75459	probable endonucle
12	76	11.0	232	2 F72383	probable endonucle
13	74.5	10.8	572	2 T52520	hypothetical prote
14	74	10.7	218	2 G86758	hypothetical prote
15	72	10.4	219	2 AF1311	DNA-(apurinic or a
16	72	10.4	219	2 AF1683	probable endonucle
17	70	10.1	214	2 B71919	probable endonucle
18	70	10.1	218	2 A64593	endonuclease III -
19	69.5	10.0	213	2 H70325	endonuclease III -
20	69	10.0	218	2 B83862	endonuclease III (
21	68.5	9.9	264	2 C72770	probable DNA-(apur
22	68	9.8	219	2 D89923	endonuclease-like
23	67	9.7	209	2 C69461	endonuclease III (
24	64	9.2	219	2 I40525	endonuclease III (
25	63.5	9.2	212	2 F71634	endonuclease III (
26	62.5	9.0	225	2 B75537	endonuclease III -
27	62	8.9	297	2 C87149	probable DNA glyco
28	61.5	8.9	292	2 B97233	8-oxoguanine-DNA-g
29	61	8.8	209	2 F95148	endonuclease III (

30 61 8.8 209 -2 D98016 DNA-(apurinic or a  
31 60.5 8.7 211 2 G96984 probable endonucle  
32 59 8.5 350 2 B38535 A/G-specific adeni  
33 59 8.5 350 2 E91108 adenine glycosylas  
34 59 8.5 350 2 H85953 adenine glycosylas  
35 58.5 8.4 220 2 A64479 DNA-(apurinic or a  
36 57.5 8.3 142 2 A47142 cystatin D precurs  
37 57.5 8.3 218 2 E82780 endonuclease III X  
38 56.5 8.2 212 2 H83208 endonuclease III P  
39 56.5 8.2 213 2 F72387 endonuclease III -  
40 56.5 8.2 220 2 D70316 endonuclease III -  
41 56.5 8.2 236 2 AE3104 endonuclease III ( -  
42 56.5 8.2 260 2 F98182 endonuclease III P  
43 56.5 8.2 376 2 S49801 OGG1 protein - yea  
44 56 8.1 208 2 B81407 endonuclease III C  
45 56 8.1 355 2 G83003 A / G specific ade

## ALIGNMENTS

### RESULT 1

T48453

hypothetical protein T32M21.170 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48453

R:Bevan, M.; Tarryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 2000

A:Reference number: 224487

A:Accession: T48453

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-555 <BEV>

A:Cross-references: EMBL:AL162875

A:Experimental source: cultivar Columbia; BAC clone T32M21

C:Genetics:

A:Map position: 5

A:Introns: 469/3; 496/2; 524/3

A>Note: T32M21.170

Query Match 32.9%; Score 228; DB 2; Length 555;

Best Local Similarity 44.5%; Pred. No. 9e-28;

Matches 57; Conservative 0; Mismatches 61; Indels 10; Gaps 1;

Oy 18 DXXXXXXXRXNX

Db 436 DSIDYEATRRASISEISEAIAKERGMNMLAVRI-----KDFLEIRIVKHGGIDLE 485

Oy 78 WLRXXXXXXXRXNX

Db 486 WLRSPDPKADYLLSIRGLKSKVECVRLTLHLNLAFFPDTNVGRIVRXXVPLXPLP 137

Oy 138 XXXQXHL 145

Db 546 ESLQLHL 553

### RESULT 2

T05430

hypothetical protein F28A23.180 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 24-Nov-1999

C:Accession: T05430

R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.;

submitted to the Protein Sequence Database, October 1998

A:Reference number: 215415

A:Accession: T05430

A:Molecule type: DNA

A:Residues: 1-917 <BEV>

A:Cross-references: EMBL:AL021961

A:Experimental source: cultivar Columbia; BAC clone F28A23

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (uspto)**



A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: D84198  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-265 <STO>  
A:Cross-references: GB:AE004437; NID:g10580010; PIDN:AAG18952.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: nthA1

Query Match 12.9%; Score 89.5; DB 2; Length 265;  
Best Local Similarity 27.6%; Pred. No. 3.9e-06;  
Matches 32; Conservative 8; Mismatches 53; Indels 23; Gaps 5;  
QY 88 KXLLXXGXGLKSECVRLRLXX-XAFPVDTNVRGXVPLXPLPXXXGXHXLX 146  
DB 145 RAALLDMTGVGPKTADCVLLFAGGAGVFPVDTVHRTARRMGIAPAADAEVRAALEA 204  
QY 147 YXPXXXXXOKLWPLRCLKXQXTLYELHYXXITFGKXXFXCTKXXPNC-----NACPM 198  
DB 205 AVPDAA-----CGFG-----HTAMIQFGR-EYCTARDPACLDDPAACPL 242

RESULT 7  
GB7530  
endonuclease III family protein [Imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: GB7530  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: GB7530  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-241 <STO>  
A:Cross-references: GB:AE005673; NID:g13423785; PIDN:AAK24243.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC2272

Query Match 12.5%; Score 86.5; DB 2; Length 241;  
Best Local Similarity 24.2%; Pred. No. 1.1e-05;  
Matches 31; Conservative 9; Mismatches 73; Indels 15; Gaps 3;  
QY 72 GXIDLEWLRXXXXXXXKXLLXXGXGLKSECV-RLRLXXXXAFPVDTNVRGXVPLX 130  
DB 106 GWLSLSHLTKTLEVDQARWELQALPGVGVKVAACVLFNFSDLAMRALVDTVDRVARRIGL 165  
QY 131 VPLXPLPXXXQXHLXLYXPXXXXXOKLWPLRCLKXQXTLYELHYXXITFGKXXFXCTKXX 190  
DB 166 V-----GSGDNTTYHTLMAPASW-----TADDLFELHLMKRGQLGQMLCGAEA 211  
QY 191 PNCNACPM 198  
DB 212 PKCGACPV 219

RESULT 8  
A75109  
endonuclease III (nth) PAB0459 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: A75109  
R:Anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome stru  
A:Reference number: A75001  
A:Accession: A75109  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-222 <KAW>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49586.1; PID:g545  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: nth; PAB0459  
C:Superfamily: apurinic/apyrimidinic endonuclease III  
C:Keywords: 4Fe-4S; metalloprotein  
F:195,202,205,211/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 12.4%; Score 86; DB 2; Length 222;  
Best Local Similarity 26.9%; Pred. No. 1.2e-05;  
Matches 29; Conservative 11; Mismatches 50; Indels 18; Gaps 4;  
QY 91 LLXXXXGXGLKSECVRLRLXXXXAFPVDTNVRGXVPLXPLPXXXGXHXLXLYPX 150  
DB 118 LKLPFGIGRKCANIVLAYGFGIPAPVDTVHVRISRRGLAPMDASPEVEERL----- 171  
QY 151 XXXXOKXLMWPLRCLKXQXTLYELHYXXITFGKXXFXCTKXXPNCNACPM 198  
DB 172 -----KELIPR-----EENIY-VNHAMVDHGK-SVCRPIKPRCDECP 207

RESULT 9  
B69202  
endonuclease III - Methanobacterium thermoautotrophicum (strain Delta H)  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 16-Jul-1999  
C:Accession: B69202  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T  
.; Qiu, D.; Spadafora, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanl,  
K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu  
A:Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: B69202  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-233 <MTH>  
A:Cross-references: GB:AE000855; GB:AE000666; NID:g2621852; PIDN:AAB85267.1; PID:g262  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH764  
A:Start codon: TTG  
C:Superfamily: apurinic/apyrimidinic endonuclease III

Query Match 12.0%; Score 83.5; DB 2; Length 233;  
Best Local Similarity 26.9%; Pred. No. 3.1e-05;  
Matches 29; Conservative 10; Mismatches 50; Indels 19; Gaps 4;  
QY 91 LLXXXXGXGLKSECVRLRLXXXXAFPVDTNVRGXVPLXPLPXXXGXHXLXLYPX 150  
DB 127 LKLPFGVGRKTCANCLVLYAFGRPAIPVDTVHVRISNRIGLVD-TRTPETERALMKVIP- 184  
QY 151 XXXXOKXLMWPLRCLKXQXTLYELHYXXITFGKXXFXCTKXXPNCNACPM 198  
DB 185 -----REYW-----IELNDLMVQFG-DICRPLGPRHEECP 215

RESULT 10  
F71025  
probable endonuclease III - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C:Accession: F71025  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, T.; Yamazaki, J.; Kushida, N.; Ogu  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: F71025  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-222 <KAW>

**THIS PAGE BLANK (uspto)**

A:Cross-references: GB:AP000006; NID:g23236133; PIDN:BAA30606.1; PID:g93257923  
A:Experimental source: strain OT3  
A>Note: This accession replaces an interim accession for a sequence replaced by GenBank C:Genetics:  
A:Gene: PH1498  
C:Superfamily: apurinic/aprimidinic endonuclease III  
C:Keywords: 4Fe-4S; iron-sulfur protein; metalloprotein  
F:195,202,205,211/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match            11.88; Score 82; DB 2; Length 222;  
Best Local Similarity 26.99; Pred. No. 5.1e-05;  
Matches 29; Conservative 10; Mismatches 51; Indels 18; Gaps 4;

Qy 91 LLXXGXGKLGXECVRLILXXXAFPVDVTNGRIXRGXVPLXPPLPXXXXHXLXXYPX 150  
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 118 LKLPGIGRKCANIVAYGFGIPAIPDVTHYRISRRLLAPWDASPEEVEERL----- 171  
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 151 XXXXOKLWLPRLCLXONTLYELHYXXITFGKXXFCFKXXPNACPM 198  
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 172 -----KSLIPR-----EEWII-VNHAMVDHGK-SVCRPKPRCWECPL 207  
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11  
C75459  
probable endonuclease III - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: C75459  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M., M.; Shen, M.; VamaChevan, J.J.; Lam, P.; McDonald, T.; Zalewski, C.; Fraser,  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75459  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-338 <WHI>  
A:Cross-references: GB:AE001945; GB:AF000513; NID:g6458642; PIDN:AAF10505.1; PID:g645865  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR0928  
A:Map position: 1

[illegible]

RESULT 12  
F72383  
hypothetical protein TW0382 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: F72383  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; PMID:99287316; PMID:10360571

A:Accession: F72383  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-232 <ARN>  
A:Cross-references: GB:AE001718; GB:AF000512; NID:94980881; PIDN:AAD35467.1; PID:9498  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0382  
C:Superfamily: Methanococcus jannaschii conserved DNA-(apurinic or apyrimidinic site)

Query Match	11.0%	Score 76;	DB 2;	Length 232;
Best Local Similarity	22.5%;	Pred. No. 0.00047;		
Matches 29;	Conservative 14;	Mismatches 68;	Indels 18;	Gaps 3;
Qy	75	DLEWLXXXXXXXXXXLLXXGXGLKSECVRLXLXXAFVDTNVGRIAVRXGXVPLX	134	
Db	113	NLSRLDPTHLRRLUKIGIKETADALLVALEKPFVWDSYTRRLUKRTFNELN	172	
Qy	135	PLPXXXXHXLXLYPPXXXXXOKLWPLRCLKXQTXTELHYXXITFGXXXXFCTKXXPCNC	194	
Db	173	DY-DEVQKLFWTHYP-----EDVRLVQEFHGLVHAHKFCSK-TPKCG	214	
Qy	195	ACPMXXHEC	203	
Db	215	VCPLKKECC	223	

```

RESULT 13
T52520
hypothetical protein B2J23.160 [Imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000
C:Accession: T52520
R:Schultze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakutu
submitted to the Protein Sequence Database, September 2000
A:Reference number: Z26053
A:Accession: T52520
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-572 <SCH>
A:Cross-references: EMBL:AL442164; GSPDB:GN00116; NCSP:B2J23.160
A:Experimental source: BAC clone B2J23; strain OR74A
C:Genetics:
A:Gene: NCSP:B2J23.160
A:Map position: 6
A:Introns: 67/1

```

Query Match	10.8%	Score 74.5;	DB 2;	Length 572;
Best Local Similarity	24.1%	Pred. No. 0.0018;		
Matches 34;	Conservative 10;	Mismatches 68;	Indels 29;	Gaps 4;
Qy	68	VXXGXGIDLEWLRLXXXXXXXLLXXXXGXGLKSXECVRLLLXXXXAFPUDTNGRLXVR	127	
		:::	:::	:::
Db	436	VAFSETUNLDHMTVTKDEAMAKILVQYPGIGISAACTVFLCLRMPCFADVTHVHRCRW	495	
		:::	:::	:::
Qy	128	YGXVPLPLPXXXXQXHLXXYPXXXXXQKLWPLRC--KLXOXTLYELHYXXITFGKXF	185	
		:::	:::	:::
Db	496	LGWTPXKANAEDCFH-----CDVKVPDHLKYGULHOLFIRHGQOCF	536	
Qy	186	-C-----TKXXPNCNACPM	198	
		:::	:::	:::
Db	537	KCKATRPGTKEWREAPCEPL	557	

RESULT 14  
G86758  
DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) [imported] - Lactococcus lac  
N:Alternate names: endonuclease III  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: G86758  
R:Bolotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh  
Genome Res. 11, 731-753, 2001

**THIS PAGE BLANK (USPTO)**

A:Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ss  
A:Reference number: A86625; MUID:21235186; PMID:1137471  
A:Accession: G86758  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-218 <STO>  
A:Cross-references: GB:AE005176; PID:g12724026; PIDN:AAK05169.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: nth  
C:Superfamily: apurinic/apryrimidinic endonuclease III  
C:Keywords: carbon-oxygen lyase

Query Match 10.7%; Score 74; DB 2; Length 218;  
Best Local Similarity 25.9%; Pred No. 0.00092;  
Matches 30; Conservative 9; Mismatches 49; Indels 28; Gaps 4;  
QY 88 KXXLLXXGXGKSECVRLXLLXXXAFVDTNVGRXVPLXPLPXXXXHXLXX 147  
DQ 112 KKVLTLPQVGRKTNVLAEAVGIPGIAVDTHVERVSKRLDIVP----- 156  
QY 148 YPXXXXXXQX-----LWPRCLKX-QXTLYELHYXXITFGKXXFCTKXXPNACPM 198  
DQ 157 -----QKATVLEVEEKLMKLIPOEKWVOAHHHLIFFGR-YHCTAKPKCADCPV 204

## RESULT 15

AF1311  
probable endonuclease III (DNA repair) [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AF1311  
R:Glaser, P.; Francœur, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.; Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.: Comparative genomics of *Listeria* species.  
A:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AF1311  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC99972.1; PID:g16411347; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: nth  
C:Superfamily: apurinic/apryrimidinic endonuclease III

Query Match 10.4%; Score 72; DB 2; Length 219;  
Best Local Similarity 25.2%; Pred No. 0.0019;  
Matches 26; Conservative 8; Mismatches 51; Indels 18; Gaps 3;  
QY 96 GXGLKSECVRLXLLXXXAFVDTNVGRXVPLXPLPXXXXHXLXXYPXXXXXQ 155  
DQ 117 GVGKRTANVLSVGFQVPAIAVDTHVERISKRLGICRWKDSVVEETLKRKL----- 170  
QY 156 KXWPRCLKXQXTLYELHYXXITFGKXXFCTKXXPNACPM 198  
DQ 171 KELWS-----DAHHYIFFGR-YHCKARNPECPTCL 201

Search completed: November 8, 2002, 19:26:43  
Job time : 17.0327 secs

**THIS PAGE BLANK (USPTO)**



**THIS PAGE BLANK (USPTO)**





	Query Match	60.5%;	Score 166.5;	DB 10;	Length 1309;
	Best Local Similarity	43.4%;	Pred. No. 3.2e-20;		
	Matches 43;	Conservative 0;	Mismatches 20;	Indels 9;	Gaps 2;
QY	1	KVXDXTXXXXXX	XXXXXXXXXXXX	-----	XXERXXFXRXRXFXFXRXXXQGX
Db	497	KVQLDPETSRV	NKLLMSS	DCDCVDCSDEE	KRWEEERNMFGHANSFARMVVOG
OV	53	FXFXMKGSVV	DSVGVGFLTON	-XDXSSSXAXXX	AXXFP 90

```
Db 882 LGLKSVCEVRLSLHQAIPVDTNNGRIAVLGGWVLPQLPDLQMLL----- 930
QY 157 XLWRLCLXQVLYELHYXITFGKXXFCTKXXPNACPMXXEAXXSA 209
Db 931 -----ELYELHYHMITFGK-VFCTKPNACPM-KAECHRYSSA 969

RESULT 2
Q91Z68 PRELIMINARY; PRT; 555 AA.
AC Q91Z68;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 62.1 kDa protein.
GN T32M21.170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardles W., Buyshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162875; CAB85563.1;
DR InterPro: IPR003265; Endo-3c.
DR Pfam: PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR Hypothetical protein.
SQ SEQUENCE 555 AA; 62088 MW; ALB44BDDE17FDC1E CRC64;

Query Match
Best Local Similarity 32.9%; Score 228; DB 10; Length 555;
Matches 57; Conservative 0; Mismatches 61; Indels 10; Gaps 1;

QY 18 DXXXXXXXXXXXXXXXXXXXXGXXXXXLLXRIXXXXXXXFLXXXXXXGIDLE 77
Db 436 DSIDYEATRRASISSEISAIRKGMNMLAVRI-----KDFLERIVKDHGIDLE 485
QY 78 WLXXXXXXXLLXGXXGLKXCEVRLXLLXXXXXAFVDTNNGRIXXGVPLXPLP 137
Db 486 WLRESPPDKADYLLSIRGLGLKSVCEVRLTLHNLAFVDTNNGRIAVRGGWVLPQLP 545

QY 138 XXXQXHL 145
Db 546 ESLQLHL 553

RESULT 3
Q94LX6 PRELIMINARY; PRT; 119 AA.
AC Q94LX6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative FPSynthase (Fragment).
GN FRPFS1.
OS Euphorbia tirucalli.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.
OX NCBI_TaxID=142860;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardles W., Buyshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villaroel R., Gielen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021961; CAAL17566.1;
DR EMBL; AL161584; CAB80123.1;
DR InterPro: IPR003265; Endo-3c.
DR InterPro: IPR003651; Fes bind.
DR Pfam: PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
DR Hypothetical protein.
SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match
Best Local Similarity 29.1%; Score 201.5; DB 10; Length 917;
Matches 46; Conservative 0; Mismatches 22; Indels 25; Gaps 3;

QY 117 VDTNNGRIXXGVPLXPLPXXQXHXLLXXYPXXXXXQXLLWRLCLXQVLYELHYX 176
Db 625 VDTNNGRIAVRLGLVPLPLEPLFNGVQHQLF-----YELHYQ 661
QY 177 XITFGKXXFCTKXXPNACPMXXEAXXSA 209
Db 662 MITFGK-VFCTKPNACPM-KSECKYFASA 692
```

```
RA Asada T., Fukusaki E., Kobayashi A., Sando T.;
RT "Prenyltransferase of Euphorbia tirucalli.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051105; BAB47587.1;
DR InterPro: IPR003265; Endo-3c.
DR Pfam: PF00730; HHH-GPD; 1.
DR NON_TER 1
FT NON_TER 119
SQ SEQUENCE 119 AA; 13771 MW; FDBCE4CD0CC5B81 CRC64;

Query Match
Best Local Similarity 29.1%; Score 202; DB 10; Length 119;
Matches 50; Conservative 0; Mismatches 39; Indels 10; Gaps 1;

QY 37 IXRGXXXXLXRIXXXXXXXFLXXXXXGIDLEWLRLXXXXXXXLXXG 96
Db 31 IKERGMNVLAKRI-----KDFLNLRLVREHGNDLEWLRLVDPDKAKEVLLSIRG 80
QY 97 XGLKXCEVRLXLLXXXXXAFVDTNNGRIXXGVPLXPLP 135
Db 81 LGLKSVCEVRLTLHNLAFVDTNNGRIAVRGGWVLPQLP 119
```

```
RESULT 4
Q49498 PRELIMINARY; PRT; 917 AA.
AC Q49498;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 106.3 kDa protein.
GN F28A23.180 OR AT4G34060.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
RA Herzi A., Neumann S., Hoheisel J., Jesse T., Heijnen L., Vos P.,
RA Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzi A.,
RA Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021961; CAAL17566.1;
DR EMBL; AL161584; CAB80123.1;
DR InterPro: IPR003265; Endo-3c.
DR InterPro: IPR003651; Fes bind.
DR Pfam: PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
DR Hypothetical protein.
SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;
```

```
Query Match
Best Local Similarity 29.1%; Score 201.5; DB 10; Length 917;
Matches 46; Conservative 0; Mismatches 22; Indels 25; Gaps 3;

QY 117 VDTNNGRIXXGVPLXPLPXXQXHXLLXXYPXXXXXQXLLWRLCLXQVLYELHYX 176
Db 625 VDTNNGRIAVRLGLVPLPLEPLFNGVQHQLF-----YELHYQ 661
QY 177 XITFGKXXFCTKXXPNACPMXXEAXXSA 209
Db 662 MITFGK-VFCTKPNACPM-KSECKYFASA 692
```

Result No.	Query			DB	ID	Description
	Score	Match	%			
1	335.5	48.4	1309	10	Q9SR66	Q9sr66 arabidopsis
2	228	32.9	555	10	Q9LZ68	Q9Lz68 arabidopsis
3	202	29.1	119	10	Q9ALX6	Q9Alx6 euphorbia t
4	201.5	29.1	917	10	Q94398	Q94398 arabidopsis
5	139	20.1	1207	10	Q9S3Q6	Q9Sjq6 arabidopsis
6	112.5	16.2	259	16	Q9RRQ0	Q9rrq0 deinococcus
7	89.5	12.9	236	17	P95945	P95945 sulfolobus
8	89.5	12.9	265	17	Q9HS56	Q9hs56 halobacteri
9	87.5	12.6	223	17	Q9P9L7	Q9p9l7 pyrobaculum
10	86.5	12.5	241	16	Q9A623	Q9a623 caulobacter
11	86	12.4	222	17	Q9V0W6	Q9v0w6 pyrococcus
12	84	12.1	225	17	Q8UIH8	Q8ulh8 pyrococcus
13	83.5	12.0	233	17	Q28658	Q28658 methanobact
14	83	12.0	224	17	Q971F2	Q971f2 sulfolobus
15	82	11.8	222	17	O59167	O59167 pyrococcus
16	78.5	11.3	338	16	Q9RVU4	Q9rvu4 deinococcus

Search completed: November 8, 2002, 19:28:34  
Job time : 14.1765 secs

```

DE Hypothetical 82.2 kDa protein.
GN T23K23.26.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., Brooks S.Y.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Chinn C.W.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC012563; AAGS2018.1; -.
DR HSSP; P12931; IFMK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001610; PAS.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PD00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR TIGRFAMS; TIGR00229; sensory_box; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Hypothetical protein; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 738 AA; 82220 MW; B2B65569A52D6145 CRC64;

Query Match 15.8%; Score 43.5; DB 10; Length 738;
Best Local Similarity 45.8%; Pred. No. 75;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Qy 58 KGSVVDSVXGVFLTONDXXSXA 81
:|::| | | | | | | | | | | | | |
Db 333 RGNLIQSPFGVFLC-NDDKSSSKA 355

RESULT 13
Q93YU0
ID Q93YU0 PRELIMINARY; PRT; 765 AA.
AC Q93YU0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 85.1 kDa protein.
GN Atlg67890.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Banth J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene Atlg67890 (GI:15220577).";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY059769; AAL24117.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00086; PAC; 1.
DR TIGRFAMS; TIGR00229; sensory_box; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN1.
KW ATP-binding; Hypothetical protein; Transferase.
SQ SEQUENCE 765 AA; 85149 MW; 6651738FB4E347D4 CRC64;

Query Match 15.8%; Score 43.5; DB 10; Length 765;
Best Local Similarity 45.8%; Pred. No. 78;
Matches 11; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

Qy 58 KGSVVDSVXGVFLTONDXXSXA 81
:|::| | | | | | | | | | | | | |
Db 333 RGNLIQSPFGVFLC-NDDKSSSKA 355

RESULT 14
Q8ZW64
ID Q8ZW64 PRELIMINARY; PRT; 120 AA.
AC Q8ZW64;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PAE1954.
GN PAE1954.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009849; AAL63838.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 120 AA; 13398 MW; 34DF06D062B6AD5E CRC64;

Query Match 15.6%; Score 43; DB 17; Length 120;
Best Local Similarity 38.9%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 54 FXWXKGSVVDSVXGVFLT 71
:|::| | | | | | | | | | | | | |

```



```

DR EMBL; AK020704; BAB32184.1; -.
DR MGD; MGI:1924824; A230020K05Rik.
SQ SEQUENCE 207 AA; 22411 MW; 558A70EC798BB4C9 CRC64;

Query Match 16.7%; Score 46; DB 11; Length 207;
Best Local Similarity 40.9%; Pred. No. 6.7;
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 52 RXFXXWKGVSDSVXGVFLTON 73 .
| : : : : : : : : : : : : : : : :
Db 165 RVLFAFPGLNLTVSAGVYFTDN 186

RESULT 7
Q9Y4C2 PRELIMINARY; PRT; 921 AA.
AC Q9Y4C2; Q9BW63;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA0738 protein.
DE KIAA0738
GN KIAA0738
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
RN [2]
RP SEQUENCE OF 1-919 FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB018281; BAA34458.1; -.
DR EMBL; BC000609; AAH00609.1; -.
SQ SEQUENCE 921 AA; 102139 MW; 03D39FDF90561F74 CRC64;

Query Match 16.7%; Score 46; DB 4; Length 921;
Best Local Similarity 33.3%; Pred. No. 35;
Matches 9; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 52 RXFXXWKGVSDSVXGVFLTONDXXS 78
| : : : : : : : : : : : : : : : :
Db 165 RVLFTFPGNLVTSVAGVYFTDNKGDTS 191

RESULT 8
Q9PPT3 PRELIMINARY; PRT; 1131 AA.
AC Q9PPT3
AC Q9PPT3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical membrane lipoprotein.
GN U0557.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OC NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SERVOAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;

```



1924 H. / NEUMANN O. / HOLTZET O. / BESSE I. / NEIJEN L. / VOS F. /

GenCore version 5.1.3

Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 11.1765 Seconds  
(without alignments)  
1659.222 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KVVXDXTXXXXLXXD.....TQNXDXXSSAXMXXAXFP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum Match 100%

Listing first 45 summaries

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

1: sp-archaea.\*

2: sp-bacteria.\*

3: sp-fungi.\*

4: sp-human.\*

5: sp-invertebrate.\*

6: sp-mammal.\*

7: sp-mhc.\*

8: sp-organelle.\*

9: sp-phage.\*

10: sp-plant.\*

11: sp-rodent.\*

12: sp-virus.\*

13: sp-vertebrate.\*

14: sp-unclassified.\*

15: sp-rvirus.\*

16: sp-bacteriap.\*

17: sp-archeap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	166.5	60.5	1309	10 Q9SR66	Q9sr66 arabidopsis
2	163.5	59.5	1207	10 Q9SJ06	Q9sj06 arabidopsis
3	162	58.9	917	10 Q49498	Q49498 arabidopsis
4	49.5	18.0	1017	10 Q9LZ69	Q9LZ69 arabidopsis
5	47	17.1	541	5 Q09617	Q09617 caenorhabd
6	46	16.7	207	11 Q9D223	Q9d223 mus musculu
7	46	16.7	921	4 Q9Y4C2	Q9y4c2 homo sapien
8	45	16.4	1131	16 Q9PPT3	Q9ppt3 ureaplasma
9	44.5	16.2	373	10 Q9FNQ3	Q9fnq3 arabidopsis
10	44	16.0	1141	3 Q9P300	Q9p3u0 schizosacch
11	43.5	15.8	327	10 Q9S260	Q9sz60 arabidopsis
12	43.5	15.8	738	10 Q9C9V5	Q9c9v5 arabidopsis
13	43.5	15.8	765	10 Q93YU0	Q93yu0 arabidopsis
14	43	15.6	120	17 Q8ZM64	Q8zm64 pyrobaculum
15	43	15.6	128	10 Q8RWM5	Q8rwm5 arabidopsis
16	43	15.6	217	10 Q9C9E0	Q9c9e0 arabidopsis

17	43	15.6	294	16 Q9K921	Q9k921 bacillus ha
18	43	15.6	306	5 Q9BK13	Q9bk13 paramecium
19	43	15.6	428	10 Q9LV81	Q9lv81 arabidopsis
20	43	15.6	434	16 Q92983	Q929e3 listeria in
21	43	15.6	434	16 Q8Y544	Q8y544 listeria mo
22	43	15.6	474	3 Q9P8D6	Q9p8d6 aspergillus
23	43	15.6	565	5 Q23010	Q23010 caenorhabd
24	43	15.6	832	16 Q9X0V8	Q9x0v8 thermotoga
25	43	15.6	1127	12 P87638	P87638 dengue viru
26	42	15.3	190	12 Q8L60	Q8l60 olive leaf
27	42	15.3	258	4 Q9BP22	Q9bp22 homo sapien
28	42	15.3	370	4 Q9NUZ1	Q9nuz1 homo sapien
29	42	15.3	414	2 Q93GK8	Q93gk8 klebsiella
30	42	15.3	437	16 Q8UEE5	Q8uee5 agrobacteri
31	42	15.3	460	10 Q9FVNO	Q9fvno lycopersico
32	42	15.3	868	16 Q9KRY9	Q9kry9 vibrio chol
33	42	15.3	894	10 Q9ZTA4	Q9zta4 arabidopsis
34	42	15.3	1062	2 Q9RC22	Q9rc22 bacillus sp
35	41.5	15.1	119	17 Q9HHX9	Q9hhx9 halobacteri
36	41.5	15.1	291	10 Q94DV2	Q94dv2 oryza sativ
37	41.5	15.1	723	5 Q17748	Q17748 caenorhabd
38	41.5	15.1	763	10 Q9C903	Q9c903 arabidopsis
39	41.5	15.1	777	10 Q9C833	Q9c833 arabidopsis
40	41	14.9	180	12 Q91RP0	Q91kp0 dengue viru
41	41	14.9	247	2 Q9XCE9	Q9xce9 streptomyce
42	41	14.9	252	16 Q9JY82	Q9jy82 neisseria m
43	41	14.9	253	17 Q26841	Q26841 methanobact
44	41	14.9	262	13 Q90WC2	Q90wg2 gallus gall
45	41	14.9	311	1 Q9P9B4	Q9p9b4 uncultured

## ALIGNMENTS

RESULT 1

Q9SR66 PRELIMINARY; PRT; 1309 AA.

AC Q9SR66; DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE T22K18.18 protein.

GN T22K18.18.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,

RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,

RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence."

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC010927; AAF04422.1;

DR InterPro; IPR003265; Endo3c.

DR InterPro; IPR003651; FeS\_bind.

DR Pfam; PF00730; HNH-GPD; 1.

DR SMART; SM00478; ENDO3c; 1.

DR SMART; SM00525; FES; 1.

SQ SEQUENCE 1309 AA; 148451 MW; 3B21E4439BA4A673 CRC64;

Query Match 60.5%; Score 166.5; DB 10; Length 1309;

Best Local Similarity 43.4%; Pred. No. 3.2e-20;

Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;

Oy 1 KVVXDXTXXXXLXXDXXXXXXX-----XXERXXFXRXRXFXRXRXRXRXRXRX 52

Db 497 KVLDPETSRVKKLLMSSIDCDGVGDSDEERKKWWEERNMFHGRANSFIARMRVVQNR 556

Oy 53 FXFXWKGSVDSVGVFLTON-XDXXSSAXMXXAXFP 90

**THIS PAGE BLANK (USPTO)**

Db 118 LIKLPGRKCANIVLAYGFGIPAPVDTHVYRISRRLGLAPWDASPEVEERL----- 171

Qy 151 XXXXOKXLWRLCKLXQXTLYELHYXXITFGKXXFCTKXXPNACPM 198

Db 172 -----KSLIPR-----EEWIY-VNHAMVDHGK-SVCKPIKPRCWECP 207

Search completed: November 8, 2002, 19:28:38  
Job time : 32.5621 secs

```
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Endonuclease III.
GN MTH764.
OS Methanobacterium thermoautotrophicum.
OC Archaea: Euryarchaeota: Methanobacteria: Methanobacteriales:
OC Methanobacteriellaceae: Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzaden R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell J. D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000855; AAB85267.1; -
DR HSSP: P20625; 2ABK
DR InterPro: IPR004036; EndoIII_HhH.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR003583; HHH_1.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00525; FES; 1.
DR SMART: SM00278; HhH; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; 1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 233 AA; 27012 MW; 10F6163AAE98395B CRC64;

Query Match 12.0%; Score 83.5; DB 17; Length 233;
Best Local Similarity 26.9%; Pred. No. 1.4e-05;
Matches 29; Conservative 10; Mismatches 50; Indels 19; Gaps 4;

OY 91 LLXXGXGLKSECVRLLLXXXXAFPDVTVNGRIVXRVXGVPPLPPLXXXQXHLXXYPX 150
DB 127 LKLFQGVGRKTCNVLVYAFGRPALPVDTHVHRISNRGLVD-TTPTEERALMKVIP- 184
OY 151 XXXXQXKLWPRCLKXQXTLYELHYXXITFGKXXFCCTKXXPCNACPM 198
DB 185 -----REYW-----IELNDLMVQFGQ-DICRPLGPRHECPI 215

RESULT 14
OY971F2
ID Q971F2 PRELIMINARY; PRT; 224 AA.
AC Q971F2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2002 (TReMBLrel. 20, Last annotation update)
DE Putative endonuclease III.
GN ST1401.
OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
```

```
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000986; BAB66468.1; -
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003583; HHH_1.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00278; HhH; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; UNKNOWN_1.
KW Endonuclease; Hypothetical protein; Complete proteome.
SQ SEQUENCE 224 AA; 25629 MW; E2BE3A0762135550 CRC64;

Query Match 12.0%; Score 83; DB 17; Length 224;
Best Local Similarity 25.7%; Pred. No. 1.6e-05;
Matches 27; Conservative 13; Mismatches 47; Indels 18; Gaps 3;

OY 92 LXXGXGLKSECVRLLLXXXXAFPDVTVNGRIVXRVXGVPPLPPLXXXQXHLXXYPX 151
DB 133 LTVDIGEKATADVIVNCFKCKFFPDVTHIKRVMSRLGILGSKP-----QYKEIADF--- 174
OY 152 XXXQXKLWPRCLKXQXTLYELHYXXITFGKXXFCCTKXXPCNAC 196
DB 175 -----FISSINDELELHQLLLHGRKT-CTAKKPLCDKC 209

RESULT 15
OY9167
ID O59167 PRELIMINARY; PRT; 222 AA.
AC O59167;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 222AA long hypothetical endonuclease III.
GN PH1498.
OS Pyrococcus horikoshii.
OC Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=96791194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000006; BAA30606.1; -
DR HSSP: P20625; 2ABK.
DR InterPro: IPR004035; EndoIII_FCL.
DR InterPro: IPR004036; EndoIII_HhH.
DR InterPro: IPR003265; Endo_3c.
DR InterPro: IPR003651; Fes_bind.
DR InterPro: IPR003583; HHH_1.
DR Pfam: PF00730; HhH-GPD; 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00525; FES; 1.
DR SMART: SM00278; HhH; 1.
DR PROSITE: PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE: PS01155; ENDONUCLEASE_III_2; UNKNOWN_1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 222 AA; 25553 MW; DD096F91C969DF0C CRC64;

Query Match 11.8%; Score 82; DB 17; Length 222;
Best Local Similarity 26.9%; Pred. No. 2.4e-05;
Matches 29; Conservative 10; Mismatches 51; Indels 18; Gaps 4;

OY 91 LLXXGXGLKSECVRLLLXXXXAFPDVTVNGRIVXRVXGVPPLPPLXXXQXHLXXYPX 150
DB 127 LKLFQGVGRKTCNVLVYAFGRPALPVDTHVHRISNRGLVD-TTPTEERALMKVIP- 184
OY 151 XXXXQXKLWPRCLKXQXTLYELHYXXITFGKXXFCCTKXXPCNACPM 198
DB 185 -----REYW-----IELNDLMVQFGQ-DICRPLGPRHECPI 215
```

Db. 162 EYDELSRW-FMERLPKDYDLFHLKLIQFGR-DVCRARNPKGQCPI-GAKCPSFKSA 217

RESULT 10

Q9A623 PRELIMINARY; PRT; 241 AA.

ID Q9A623

AC Q9A623; (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Endonuclease III family protein.

GN CC2272.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter.

OX NCBI\_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 19089 / CB15;

RX MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,

RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,

RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,

RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,

RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RT "Complete genome sequence of Caulobacter crescentus.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

DR EMBL; AE005897; AAK24243.1; -

DR TIGR; CC2272; -

DR InterPro; IPR003265; Endo\_3c.

DR InterPro; IPR003651; Fes\_bind.

DR Pfam; PF00730; HHH-GPD; 1.

DR SMART; SM00478; ENDO3c; 1.

DR SMART; SM00525; FES; 1.

KW Complete proteome.

SQ SEQUENCE 241 AA; 26421 MW; EB43FBA976736F1C CRC64;

Query Match 12.5%; Score 86.5; DB 16; Length 241;

Best Local Similarity 24.2%; Pred. No. 4.5e-06;

Matches 31; Conservative 9; Mismatches 73; Indels 15; Gaps 3;

Qy 72 GXIDLEWRXXXXXXGLKSECV-RLLLXXXAFPVDVTNNGRIYVRGX 130

Db 106 GWLSHLKLEVDQARWELQALPGVGKVAACVLFSDLAMRALVVDTHVDRVARRIGL 165

Qy 131 VPLXPLPXXXXXHXLPXXXXXKXLPRLCKLXQTLVELHYXXITFGKXXFCTKXX 190

Db 166 V-----GSGDNTTNTHTLAMAPASW-----TADDLFELHMLKRGLOMTCGAEA 211

Qy 191 PNCNACPM 198

Db 212 PKCGACP 219

RESULT 11

Q9V0W6 PRELIMINARY; PRT; 222 AA.

ID Q9V0W6

AC Q9V0W6; (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Endonuclease III (NTH).

GN NTH OR PAB0459.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI\_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ORSAY;

RA Heilig R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ248285; CAB49586.1; -

DR HSSP; P20625; 2ABK.

DR InterPro; IPR004035; EndoIII\_FCL.

DR InterPro; IPR004036; EndoIII\_HhH.

DR InterPro; IPR003265; Endo\_3c.

DR InterPro; IPR003651; Fes\_bind.

DR InterPro; IPR003583; HHH\_1.

DR Pfam; PF00730; HHH-GPD; 1.

DR SMART; SM00478; ENDO3c; 1.

DR SMART; SM00525; FES; 1.

DR SMART; SM00278; HhH; 1.

DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; 1.

DR PROSITE; PS01155; ENDONUCLEASE\_III\_2; UNKNOWN\_1.

KW Endonuclease; Complete proteome.

SQ SEQUENCE 222 AA; 25541 MW; 27F0B5FC83C0DC6 CRC64;

Query Match 12.4%; Score 86; DB 17; Length 222;

Best Local Similarity 26.9%; Pred. No. 5.1e-06;

Matches 29; Conservative 11; Mismatches 50; Indels 18; Gaps 4;

Qy 91 LXXXGXGLKSECVRLLLXXXAFPVDVTNNGRIYVRGXVPLXPLPXXXXXHXLPX 150

Db 118 LKLPFGIGRKCANIVLAYGFGIPAIPTVTHYVIRSRRLGLAPWDASPEVEERL----- 171

Qy 151 XXXXOKLWPRCLKXQTLVELHYXXITFGKXXFCTKXXPNCNACPM 198

Db 172 -----KELIPR-----EEWIIY-VNHAMVDHGK-SVCRPIKPRCDECP 207

RESULT 12

Q8UIH8 PRELIMINARY; PRT; 225 AA.

ID Q8UIH8

AC Q8UIH8; (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Glycosylase putative, mutV-nth family.

GN PF1229.

OS Pyrococcus furiosus.

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

OC Pyrococcus.

OX NCBI\_TaxID=2261;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VCI / DSM 3638 / ATCC 43587 / JCM 8422;

RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;

RT "The complete sequence of the Pyrococcus furiosus genome.";

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE010230; AAL81353.1; -

KW Complete proteome.

SQ SEQUENCE 225 AA; 26157 MW; DA48762AC33BE4AE CRC64;

Query Match 12.1%; Score 84; DB 17; Length 225;

Best Local Similarity 25.9%; Pred. No. 1.1e-05;

Matches 28; Conservative 11; Mismatches 51; Indels 18; Gaps 4;

Qy 91 LXXXGXGLKSECVRLLLXXXAFPVDVTNNGRIYVRGXVPLXPLPXXXXXHXLPX 150

Db 127 LMKLPFGIGRKCANIVLAYGFGKPAIPVDTHYVIRSRRLGLAPINSTPEKVE----- 177

Qy 151 XXXXOKLWPRCLKXQTLVELHYXXITFGKXXFCTKXXPNCNACPM 198

Db 178 --EILKTLIP-----VEEWIIY-VNHAMVDHGK-SICRPIKPKCELCP 216

RESULT 13

O26858 PRELIMINARY; PRT; 233 AA.

ID O26858

AC O26858;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

KW	Complete proteome.	SQ	SEQUENCE	265 AA;	27481 MW;	CEFBIA2A1DB9735E	CRC64;	
	Query Match		12.9%;	Score 89.5;	DB 17;	Length 265;		
	Best Local Similarity		27.6%;	Pred. No. 1.6e-06;				
	Matches	32;	Conservative	8;	Mismatches	53;	Indels 23; Gaps 5;	
QY	88	KXLLXXGXGLKXSECVRLRLXAX-XAFPVDVTNVRGXVVRXGVPLXPLPXXXXXHXIX 146	:				:	
Db	145	RAALLDVTGVPKTDACVLLFAGGAGVFPVDTHVRIARMGIAPAAADHEAVRAALEA 204	:				:	
QY	147	YXPXXXXXQXWLPRCLKXQXTLYELHYXXITFGKXXFXCTKXXPNC-----NACPM 198	:				:	
Db	205	AVPDAA-----CGFG-----HTAMIQFGR-EYCTARDPACLDDPAACPL 242.	:				:	
RESULT 9								
Q9P9L7		PRELIMINARY;				223 AA.		
ID	Q9P9L7	PRELIMINARY;				223 AA.		
AC	Q9P9L7							
DT	01-OCT-2000 (Tremblrel. 15, Created)							
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)							
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)							
DE	Putative DNA glycosylase (DNA-(apurinic or apyrimidinic site) lyase)							
DE	(endonuclease III, Panch).							
GN	PAE0880.							
OS	Pyrobaculum aerophilum.							
OC	Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;							
OC	Thermoproteaceae; Pyrobaculum.							
OX	NCBI_TaxID=13773;							
OX	[1]							
RP	SEQUENCE FROM N.A.							
RP	MEDLINE=20138151; PubMed=10671447;							
RA	Yang H., Fitz-Gibbon S., Marcotte E.M., Tai J.H., Hyman E.C.,							
RA	Miller J.H.;							
RT	"Characterization of a thermostable DNA glycosylase specific for U/G							
RT	and T/G mismatches from the hyperthermophilic archaeon Pyrobaculum							
RT	aerophilum.";							
RL	J. Bacteriol. 182:1272-1279(2000).							
RP	[2]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=IM2 / ATCC 51768 / DSM 7523;							
RX	PubMed=11792869;							
RA	Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,							
RA	Miller J.H.;							
RT	"Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum							
RT	aerophilum.";							
RL	Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).							
DR	EMBL; AF222334; AAF37269.1; -							
DR	EMBL; AE009791; AAL63095.1; -							
DR	HSSP; P20625; 2ABK.							
DR	InterPro; IPR004035; EndoIII_FCL.							
DR	InterPro; IPR003265; Endo_3c.							
DR	InterPro; IPR003651; FeS_bind.							
DR	InterPro; IPR003583; HHH_1.							
DR	Pfam; PF007730; Hhh-GPD; 1.							
DR	SMART; SM00478; ENDO3c; 1.							
DR	SMART; SM00525; FES; 1.							
DR	SMART; SM00278; HHH1; 1.							
DR	PROSITE; PS00764; ENDONUCLEASE_III_1; UNKNOWN_1.							
KW	Lysase; Endonuclease; Complete proteome.							
SQ	SEQUENCE	223 AA;	25323 MW;	8A8445E9462167E3	CRC64;			
	Query Match		12.6%;	Score 87.5;	DB 17;	Length 223;		
	Best Local Similarity		26.9%;	Pred. No. 2.8e-06;				
	Matches	32;	Conservative	13;	Mismatches	51;	Indels 23; Gaps 5;	
QY	91	LLXXGXGLKXSECVRLRLXAXXAFPVDVTNVRGXVVRXGVPLXPLPXXXXXHXIXXPX 150	:				:	
Db	122	LLSLPGVGKKTADV-VLVNGLPLPAFPVDTHTRAKRWGI-----G 161	:				:	
QY	151	XXXXGXGLKXSECVRLRLXQXTLYELHYXXITFGKXXFXCTKXXPNCACPMXXXXXSA 209	:				:	

```
RESULT 5
Q9SJO6
ID Q9SJO6 PRELIMINARY; PRT; 1207 AA.
AC Q9SJO6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AT2G36490 protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Rounsley S.D., Barnstead M.E., Feldblum T.V.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006919; AAD24633.1; -.
DR InterPro; IPR003651; Fes_bind.
DR SMART; SM00525; FES; 1.
SQ SEQUENCE 1207 AA; 135916 MW; AADB6C9EE94DDF67 CRC64;

Query Match 20.1%; Score 139; DB 10; Length 1207;
Best Local Similarity 25.1%; Pred. No. 3.4e-14;
Matches 50; Conservative 0; Mismatches 65; Indels 84; Gaps 5;

QY 11 EXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 70
Db 846 EXTRSTMDVDMKAFRAADVKEVAETIKSRGMNHKLAERIQVLTLMNKTMOGFLDLVND 905
QY 71 XGIDLEWLRLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 130
Db 906 HGSIDLEWLRLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 916
QY 131 VPLXPLPXXXXXHXLLXYPXXXXXKXLMPLRCKLXQXTLYELHYXITFTGXXXXCTKXX 190
Db 917 VP-----PDKAK-----YELHYQMIFGK-VFCTKSK 942
QY 191 PNCNACPMXXXXXKXXXXXSA 209
Db 943 PNCNACPM-KGECRHFASA 960

RESULT 6
Q9RR00
ID Q9RR00 PRELIMINARY; PRT; 259 AA.
AC Q9RR00;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Endonuclease III.
GN DR2438
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcaceae.

Query Match 16.2%; Score 112.5; DB 16; Length 259;
Best Local Similarity 25.2%; Pred. No. 2.1e-10;
Matches 35; Conservative 10; Mismatches 49; Indels 45; Gaps 3;

OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R1;
RA MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1";
RL Science 286:1571-1577(1999).
DR EMBL; AE002073; AAF11977.1; -.
DR TIGR; DR2438; -.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; Fes_bind.
DR Pfam; PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; FES; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
KW Endonuclease; Complete proteome.
SQ SEQUENCE 259 AA; 28946 MW; AA07354719093547 CRC64;

Query Match 16.2%; Score 112.5; DB 16; Length 259;
Best Local Similarity 25.2%; Pred. No. 2.1e-10;
Matches 35; Conservative 10; Mismatches 49; Indels 45; Gaps 3;

QY 72 GXIDLEWLRLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 131
Db 120 GGYDLDFLDLPVKDALKWLTDLPGVGKVTASLVLLFNYPVPVDTTHRVSTRVGI 179
QY 132 P-----LXPLPXXXXXHXLLXYPXXXXXKXLMPLRCKLXQXTLYELHYXX 177
Db 180 PRMGEQAHRALLALLPPDP--LXPLPXXXXXHXLLXYPXXXXXKXLMPLRCKLXQXTLYELHYXX 209
QY 178 ITFGKXKXCTKXXPCNAC 196
Db 210 LSHGR-QVCTWTRPKGKC 227

RESULT 7
P95945
ID P95945 PRELIMINARY; PRT; 236 AA.
AC P95945;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Endonuclease III (DNA endonuclease III, probable) (NTH-1)
DE (EC 4.2.99.18).
GN NTH-1 OR SSO0116 OR C04006.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
RA Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA Doolittle W.F., Ragan M.A., Charlebois R.L.;
RA "Organizational characteristics and information content of an archaeal
RT genome: 156 kb of sequence from Sulfolobus solfataricus P2.";
RL Mol. Microbiol. 22:175-191(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 28.5621 Seconds  
(without alignments)  
1659.222 Million cell updates/sec

Title: US-09-840-743-72

Perfect score: 693

Sequence: 1 WXXXXXXXEXXXXXXX.....XXXXXXXXXXXXXXXXXXXX 230

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTRMBL21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match %	Length DB ID	Description	
1	335.5	48.4	1309 10 Q9SR66	Q9sr66 arabidopsis	
2	228	32.9	555 10 Q9L268	Q9L268 arabidopsis	
3	202	29.1	119 10 Q94LX6	Q94LX6 euphorbia t	
4	201.5	29.1	917 10 Q49498	Q49498 arabidopsis	
5	139	20.1	1207 10 Q9S0Q6	Q9S0Q6 arabidopsis	
6	112.5	16.2	259 16 Q9RR00	Q9S0Q6 arabidopsis	
7	89.5	12.9	236 17 P95945	P95945 sulfolobus	
8	89.5	12.9	265 17 Q9HS56	Q9hs56 halobacteri	
9	87.5	12.6	223 17 Q9P9L7	Q9p9l7 pyrobaculum	
10	86.5	12.5	241 16 Q9A623	Q9a623 caulobacter	
11	86	12.4	222 17 Q9V0W6	Q9v0w6 pyrococcus	
12	84	12.1	225 17 Q8UIH8	Q8uih8 pyrococcus	
13	83.5	12.0	233 17 Q26858	Q26858 methanobact	
14	83	12.0	224 17 Q971F2	Q971f2 sulfolobus	
15	82	11.8	222 17 Q59167	Q59167 pyrococcus	
16	78.5	11.3	338 16 Q9RVU4	Q9rvu4 deinococcus	

## ALIGNMENTS

## RESULT 1

Q9SR66 PRELIMINARY; PRT; 1309 AA.

ID Q9SR66

AC Q9SR66;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE T22K18.18 protein.

GN T22K18.18.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,

RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,

RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC010927; AAF04422.1;

DR InterPro; IPR003265; Endo.3c.

DR InterPro; IPR003651; Fes.Bind.

DR Pfam; PF00730; HHH-GPD; I.

DR SMART; SM00478; ENDO3c; I.

DR SMART; SM00525; FES; I.

SQ SEQUENCE 1309 AA; 148451 MW; 3B21E4439BA4G73 CRC64;

Query Match 48.4%; Score 335.5; DB 10; Length 1309;

Best Local Similarity 47.4%; Pred No. 3.1e-47;

Matches 82; Conservative 0; Mismatches 56; Indels 35; Gaps 4;

Qy 37 IXRGXXXXXXRXIXXXXXXXXFLXXXXXIDLEWLXXXXXXXLXXXXX 96

Db 832 IIRKGMNNMLAERI-----KAFNLRLVKKHGSDLEWLDPVDPKAKKEYLLSING 881

Qy 97 XGLKSXECVRLLLXXXXXAFVDNTVGRITXVRXGVPLPLPXXXXXHXLLXPXXXXXQK 156

C:Superfamily: hypothetical protein HI0107

Query Match 15.6%; Score 43; DB 2; Length 434;  
Best Local Similarity 57.1%; Pred. No. 26;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 62 VDSVXGVFLTONXD 75  
|||||  
DB 386 VDTGGWFLTONYE 399

RESULT 12

AI1723

hypothetical protein homolog lin2334 [imported] - *Listeria innocua* (strain Clip11262)

C:Species: *Listeria innocua*

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002

C:Accession: AI1723

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluteter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AI1723

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-434 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97561.1; PID:g16414857; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2334

C:Superfamily: hypothetical protein HI0107

Query Match 15.6%; Score 43; DB 2; Length 434;  
Best Local Similarity 57.1%; Pred. No. 26;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 62 VDSVXGVFLTONXD 75  
|||||  
DB 386 VDTGGWFLTONYE 399

RESULT 13

T29718

hypothetical protein K11G9.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000

C:Accession: T29718

R:Greco, T.; Bradshaw, H.

submitted to the EMBL Data Library, July 1996

A:Description: The sequence of *C. elegans* cosmid K11G9.

A:Reference number: 220672

A:Accession: T29718

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-565 <GR>

A:Cross-references: EMBL:U64853; PIDN:AAB04976.1; GSPDB:GN00023; CESP:K11G9.1

A:Experimental source: strain Bristol N2; clone K11G9

C:Genetics:

A:Gene: CESP:K11G9.1

A:Map position: 5

A:Introns: 31/3; 114/3; 144/2; 282/1; 352/1; 460/1

C:Superfamily: cholinesterase; cholinesterase homology

Query Match 15.6%; Score 43; DB 2; Length 565;  
Best Local Similarity 30.8%; Pred. No. 34;  
Matches 12; Conservative 4; Mismatches 17; Indels 6; Gaps 1;

QY 57 WKGSVVDSXGV-----FLTONDXSSXAXMXXYF 89  
|||||  
DB 374 YKEDVVDVGIRKEIFEYTKNISEGDEEAIMRKAEEF 412

RESULT 14

H72278

alpha-mannosidase-related protein - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: H72278

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: H72278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-832 <ARN>

A:Cross-references: GB:AE001779; GB:AE000512; NID:g4981777; PIDN:AAD36306.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1231

Query Match 15.6%; Score 43; DB 2; Length 832;

Best Local Similarity 64.3%; Pred. No. 52;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 57 WKGSVVDSXGVFL 70  
|||||

DB 160 WKSGDGTAVQGVFL 173  
|||||

RESULT 15

AE2799

HLVD family secretion protein [imported] - *Agrobacterium tumefaciens* (strain C58, Dup

C:Species: *Agrobacterium tumefaciens*

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AE2799

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AE2799

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-432 <KUR>

A:Cross-references: GB:AE008686; PIDN:AAL42811.1; PID:g17740257; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atul815

A:Map position: circular chromosome

Query Match 15.3%; Score 42; DB 2; Length 432;

Best Local Similarity 31.0%; Pred. No. 38;

Matches 9; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 58 KGSVVDSXGVFLTONDXSSXAXMXXYA 86  
| : | : | | | | : | : |

DB 212 KAKIVSPIDGVILTRSDPGATVASSLNA 240  
| : | : | | | | : | : |

Search completed: November 8, 2002, 19:26:41

Job time : 8.88235 secs

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick,A.M.; Sun, H.; Tall  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: C96748  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-217 <STO>  
A;Cross-references: GB:AE005173; NID:g6730759; PIDN:AAF27148.1; GSPDB:GN00141  
C:Genetics:  
A;Gene: T10D10.11  
A;Map position: 1

	Query Match	15.6%	Score 43;	DB 2:	Length 217;
Best Local Similarity	26.1%;	Pred. No. 12;			
Matches	12;	Conservative	5;	Mismatches	21;
				Indels	8;
Gaps					1;

QY 52 RXFXWKXGSVDS-----VXGVFLTQNXXSSAXMXXXXF 89  
I III::I I I I I  
Db 159 RYLPTWKNVIDAEMENPGHVVGMSLSVNAQGFGFIGAKSGAGDF 204

RESULT 10  
G83962  
hypothetical protein BH2503 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C;Accession: G83962  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A;Reference number: AB13650; MUID:20512582; PMID:11058132  
A;Accession: G83962  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-294 <STO>  
A;Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BAB06222.1; GSPDB:G  
A;Experimental source: strain C-125  
C:Genetics:  
A;Gene: BH2503  
C:Superfamily: conserved hypothetical protein HI1714

	Query Match	15.6%	Score 43;	DB 2:	Length 294;
Best Local Similarity	38.9%;	Pred. No. 17;			
Matches	7;	Conservative	5;	Mismatches	6;
				Indels	0;
Gaps					0;

QY 58 KGSVDVSXGVFLTQNXD 75  
II::I I I I I  
Db 3 KGTIVKALSGFYVQVED 20

RESULT 11  
AH1353  
hypothetical protein lmo2232 [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 17-May-2002  
R;Glaser, P.; Franguel, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AH1353  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-434 <GLA>  
A;Cross-references: GB:NC\_003210; PTDN:CAD000310.1; PID:g16411702; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C:Genetics:  
A;Gene: lmo2232



Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	163.5	59.5	1207	2	D84781	hypothetical prote	
2	162	58.9	917	2	T05430	hypothetical prote	
3	49.5	18.0	1017	2	T8452	conserved hypothet	
4	47	17.1	442	2	E6958	hypothetical prote	
5	47	17.1	541	2	T28069	hypothetical prote	
6	45	16.4	1131	2	F82875	hypothetical prote	
7	43.5	15.8	327	2	T06602	hypothetical prote	
8	43.5	15.8	738	2	F96701	hypothetical prote	
9	43	15.6	217	2	C96748	unknown protein T1	
10	43	15.6	294	2	G83962	hypothetical prote	
11	43	15.6	434	2	AH1353	hypothetical prote	
12	43	15.6	434	2	AI1723	hypothetical prote	
13	43	15.6	565	2	T29718	hypothetical prote	
14	43	15.6	832	2	H72278	alpha-mannosidase-	
15	42	15.3	432	2	AE2799	HLyD family secret	
16	42	15.3	437	2	G97578	probable periplasm	
17	42	15.3	868	2	G82193	aminopeptidase N V	
18	42	15.3	894	2	T01385	probable phosphor	
19	41.5	15.1	723	2	T30094	hypothetical prote	
20	41	14.9	108	1	WRBP69	early protein gp13	
21	41	14.9	108	1	WRBP13	early protein gp13	
22	41	14.9	108	2	JN0031	early protein gp13	
23	41	14.9	252	2	D81052	acylttransferase, p	
24	41	14.9	253	2	G69199	endonuclease III-r	
25	41	14.9	260	2	H69874	conserved hypothet	
26	41	14.9	280	2	D84968	hypothetical prote	
27	41	14.9	319	2	T35933	hypothetical prote	
28	41	14.9	444	2	F69826	hemolysin homolog	
29	41	14.9	638	2	T24676	hypothetical prote	

**THIS PAGE BLANK (USPTO)**

[illegible]





CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.

XX SQ Sequence 145 AA;  
Query Match 20.3%; Score 141; DB 23; Length 145;  
Best Local Similarity 45.1%; Pred. No. 2.8e-15;  
Matches 37; Conservative 0; Mismatches 35; Indels 10; Gaps 1;  
QY 37 IXRGXXXXLXXRXXXXXXXFLXXXXXGXIDLEWLRXXXXXXXKXLLXXG 96  
DB 74 IIRGMNVLSRI-----QEFNLVRDHGSIDLEWLRDIPPSAKDVLLSIRG 123  
QY 97 XGLKSECEVRLLLXXAAPPVD 118  
DB 124 LGLKSECEVRLLLTLLHLAAPPVD 145

RESULT 11  
AAU72747  
ID AAU72747 standard; Protein; 276 AA.  
XX AC  
XX AAU72747;  
XX 26-FEB-2002 (first entry)  
XX DE Corn DMT3 protein.  
XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
XX KW DNA demethylation; transgenic plant; transcription modulation;  
XX KW flowering time; endosperm development; MEDEA.  
XX OS Zea mays.

XX PN WO200180626-A1.  
XX PD 01-NOV-2001.  
XX PF 23-APR-2001; 2001WO-US13059.  
XX PR 21-APR-2000; 2000US-0553690.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX DR WPI; 2002-055307/07.  
XX PS Disclosure; Page 86; 109pp; English.  
XX The invention relates to an isolated polynucleotide sequence or their  
XX complement encoding a polypeptide having a sequence at least 40%  
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
XX or C or their combinations. Also included are an expression  
XX cassette comprising the polynucleotide or comprising a heterologous  
XX polynucleotide under the control of a promoter at least 70% identical to  
XX DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
XX region of DMT, a host cell comprising an exogenous polynucleotide  
XX encoding a DMT-like protein and a transgenic plant comprising a  
XX polynucleotide encoding a DMT-like protein. The expression cassette is  
XX useful for modulating transcription. The method comprises introducing  
XX the cassette into a host cell preferably Agrobacterium by sexual  
XX cross, and selecting a host cell with modulated transcription, where  
XX the protein is capable of exhibiting at least one of the following  
XX biological activities, which include enhanced expression of the  
XX protein in a plant results in a delay in flowering time, introduction of  
XX the protein into a cell results in modulation of methylation of

CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expression of the  
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.

XX SQ Sequence 276 AA;  
Query Match 15.3%; Score 106; DB 23; Length 276;  
Best Local Similarity 27.3%; Pred. No. 4.2e-09;  
Matches 38; Conservative 12; Mismatches 65; Indels 24; Gaps 3;  
QY 72 GXIDLEWLRXXXXXXXKXLLXXGXGLKSECEVRLLLXXAAPPVDNVRGIXVRGXV 131  
DB 147 GKICLEYLRELSVDEVKELSRFKGIGKTVACVLMFYLOKDDFVDTHTVLRITKAMGW 206  
QY 132 PLXPLPXXXXXHXLLXXYPXXXXXKXLMWRCLKXQXTLYELHYXXITFGKK-XFCT--- 187  
DB 207 PATASREKAYIHLNN-----KIPDDLKFDLNCFLFVTHGKLCQSCCTKKV 249  
QY 188 ---KXXPNCNACPMXXECC 203  
DB 250 GSKRKSNSACPLAGYCC 268

RESULT 12  
AAU72746  
ID AAU72746 standard; Protein; 166 AA.  
XX AC  
XX AAU72746;  
XX 26-FEB-2002 (first entry)  
XX DE Corn DMT2 protein.  
XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
XX KW DNA demethylation; transgenic plant; transcription modulation;  
XX KW flowering time; endosperm development; MEDEA.  
XX OS Zea mays.

XX PN WO200180626-A1.  
XX PD 01-NOV-2001.  
XX PF 23-APR-2001; 2001WO-US13059.  
XX PR 21-APR-2000; 2000US-0553690.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX DR WPI; 2002-055307/07.  
XX PS N-PSDB; AAS96697.  
XX PT New polynucleotide that control plant development comprising a sequence  
XX having a specific homology to DEMETER domains A, B or C -  
XX Disclosure; Page 85; 109pp; English.  
XX The invention relates to an isolated polynucleotide sequence or their  
XX complement encoding a polypeptide having a sequence at least 40%  
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
XX or C or their combinations. Also included are an expression  
XX cassette comprising the polynucleotide or comprising a heterologous  
XX polynucleotide under the control of a promoter at least 70% identical to  
XX DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
XX region of DMT, a host cell comprising an exogenous polynucleotide  
XX encoding a DMT-like protein and a transgenic plant comprising a  
XX polynucleotide encoding a DMT-like protein. The expression cassette is

CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.  
XX  
SQ Sequence 179 AA;

Query Match 36.8%; Score 255; DB 23; Length 179;  
Best Local Similarity 63.0%; Pred. No. 2.5e-34;  
Matches 51; Conservative 0; Mismatches 28; Indels 2; Gaps 2

QY 129 GXVPLPXPXQHXHKLXXYPXXXXXQLWPRCLKXQXTLYELHYXXITFGKXXFCTK 188  
| | | | | | | | | | | | | | | | | | | | | | | | |  
Db 2 GWVPLOPESLQHLLLELYPVLESIQXYLWPRCLKDQETLYELHYQMITEGK-VECTK 60  
| | | | | | | | | | | | | | | | | | | | | | | | |

QY 189 XPNACPMNXXECCXXXSA 209  
| | | | | | | | | | | | | | | | | | | | | | | | |  
Db 61 SKPNCNCPM-RAECRHFAA 80  
| | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10  
AAU72749  
ID AAU72749 standard; Protein; 145 AA.

XX  
AC AAU72749;

DT 26-FEB-2002 (first entry)

DE Wheat DMT2 protein.

XX  
XX Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
XX flowering time; endosperm development; MEDEA.  
XX

OS Triticum sp.  
XX  
XX W0200180626-A1.  
XX  
PD 01-NOV-2001.

XX  
XX 23-APR-2001; 2001WO-US13059.  
XX  
XX 21-APR-2000; 2000US-0553690.  
XX  
XX (REGC ) UNIV CALIFORNIA.

XX  
XX Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
PI  
DR NPSDB: 2002-055307/07.  
DR NPSDB: AAS96699.

XX  
XX New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A,B or C -  
PS Disclosure; Page 87; 109pp; English.  
XX

XX The invention relates to an isolated polynucleotide sequence or their  
CC complement encoding a polypeptide having a sequence at least 40%  
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
CC or C or their combinations. Also included are an expression  
CC cassette comprising the polynucleotide or comprising a heterologous  
CC polynucleotide under the control of a promoter at least 70% identical to  
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
CC region of DMT, a host cell comprising an exogenous polynucleotide  
CC encoding a DMT-like protein and a transgenic plant comprising a  
CC polynucleotide encoding a DMT-like protein. The expression cassette is  
CC useful for modulating transcription. The method comprises introducing  
CC the cassette into a host cell preferably Agrobacterium by sexual  
CC cross, and selecting a host cell with modulated transcription, where  
CC the protein is capable of exhibiting at least one of the following  
CC biological activities, which include enhanced expression of the  
CC protein in a plant results in a delay in flowering time, introduction of  
CC the protein into a cell results in modulation of methylation of  
CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expressing of the

The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transposition of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.

XX

Sequence 251 AA;

Query Match 53.9%; Score 373.5; DB 23; Length 251;  
Best Local Similarity 55.3%; Pred. No. 4.6e-54;  
Matches 84; Conservative 0; Mismatches 57; Indels 11; Gaps 2;

Qy 37 IXRGXXXXLXRRIRXXXXXXXXXXXXFLLXXXGXGIDLEWLRXXXXXXXXKXLLXXGG 96  
Db 109 IKRGMNMLAERI-----QSFLNLLVDKHGGIDLEWLDPVPPDQAKEFLLSIRG 158

Qy 97 XGLSXECVRLLLXLLAXAFPVDVTNVGRIVRXGVPLPXPXXQHXLXXYPXXXXXQK 156  
Db 159 LGLSVCVRLTLHHLAFPPDTNVGRAVRGWVLPQLPESQLHLELYPVLESIQK 218

Qy 157 XLWRPRLCKLOXTLYELHYXXITFGKXXFCYK 188  
Db 219 YLWRPRLCKLDORTLYELHYQLITFGK-VFCYK 249

RESULT 8  
ABB92260  
ID ABB92260 standard; Protein; 1309 AA.  
XX  
AC ABB92260;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Herbicidally active polypeptide SEQ ID NO 1471.  
XX  
KW Herbicidal; plant; agriculture; herbicide.  
XX  
OS Arabidopsis thaliana.  
XX  
PN W0200210210-A2.  
XX  
PD 07-FEB-2002.  
XX  
PS 28-AUG-2001; 2001WO-EP09892.  
XX  
PR 28-AUG-2001; 2001WO-EP09892.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Tietjen K, Weidler M;  
XX  
DR WPI; 2002-269010/31.  
XX  
PT Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -  
XX  
XX Claim 5; SEQ ID NO 1471; 26lpp + Sequence Listing; English.  
XX  
XX The invention relates to identifying target proteins  
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
CC aligning and comparing nucleic acid or amino acid sequences from plant  
CC with nucleic acid or amino acid sequences from non-plant organisms using  
CC suitable search parameters, where plant sequences having an E-value  
CC greater by a factor of 3 than the E-value of most similar non-plant  
CC sequences are selected. The polypeptides or nucleic acids encoding them  
CC are useful for identifying modulators. The identified modulators are  
XX useful as herbicides.  
XX  
Sequence 1309 AA;

Query Match 48.4%; Score 335.5; DB 23; Length 1309;  
Best Local Similarity 47.4%; Pred. No. 6.5e-47;

Db 1554 YLWPRLCKLDQRTLYELHYQMITEGK-VFCTKSKPNCNACPM-RAECKHFASA 1604

## RESULT 5

AAU072732  
ID AAU72752 standard; Protein; 298 AA.

XX AC AAU72752:

DT 26-FEB-2002 (first entry)

DE Soybean DMT1 protein.

Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
DNA demethylation; transgenic plant; transcription modulation;  
flowering time; endosperm development; MEDEA.

Glycine max.

XX  
PN WO200180626-A1.

XX  
PD 01-NOV-2001.XX  
PF 23-APR-2001; 2001WO-US13059.XX  
PR 21-APR-2000; 2000US-0553690.

XX - (REGC ) UNIV CALIFORNIA.

XX  
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV:

XX DR WPI; 2002-055307/07.

DR N-PSDB; AAS96702.

PT New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A,B or C - -  
PT

PS Disclosure; Page 88; 109pp; English.

The invention relates to an isolated polynucleotide sequence or their complement encoding a polypeptide having a sequence at least 40% identical to DMT (DEMEter, previously known as ATROPOS (ATR)) Domain A, B or C or their combinations. Also included are an expression cassette comprising the polynucleotide or comprising a heterologous polynucleotide under the control of a promoter at least 70% identical to DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated region of DMT, a host cell comprising an exogenous polynucleotide encoding a DMT-like protein and a transgenic plant comprising a polynucleotide encoding a DMT-like protein. The expression cassette is useful for modulating transcription. The method comprises introducing the cassette into a host cell preferably *Agrobacterium* by sexual cross, and selecting a host cell with modulated transcription, where the protein is capable of exhibiting at least one of the following biological activities, which include enhanced expression of the protein in a plant results in a delay in flowering time, introduction of the protein into a cell results in modulation of methylation of chromosomal DNA in the cell, reduction of expression of the protein in a plant results in enhanced endosperm development and expressing of the protein in an *Arabidopsis* leaf results in expression of the MEDEA gene. The polynucleotide is useful for detecting a nucleic acid in a sample. DEMETER is related to 5-methylcytosine glycosylases and regulates transcription of target genes by demethylation. The present sequence represents a DMT-like protein.

AA	Sequence	298 AA;
SQ		

Query Match	62.3%	Score 432;	DB 23;	Length 298;
Best Local Similarity	56.6%	Pred. No. 8.4e-64;		

Qy 37 IXXRGXXXXLXXRXIXXXXXXXXXXXXFLXXXVVXXXXGXIDLEWLRXXXXXXXXXXLLXXXG 96

21 IKERGMNNRLADRI-----KNFLNRLVEEHGSIDLEWLRDVPPDKAKEYLLSIRG 70



XX PS Disclosure; Page 67; 109pp; English.

XX CC The invention relates to an isolated polynucleotide sequence or their

XX CC complement encoding a polypeptide having a sequence at least 40%

XX CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B

XX CC or C or their combinations. Also included are an expression

XX CC cassette comprising the polynucleotide or comprising a heterologous

XX CC polynucleotide under the control of a promoter at least 70% identical to

XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

XX CC region of DMT, a host cell comprising an exogenous polynucleotide

XX CC encoding a DMT-like protein and a transgenic plant comprising a

XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

XX CC useful for modulating transcription. The method comprises introducing

XX CC the cassette into a host cell preferably Agrobacterium by sexual

XX CC cross, and selecting a host cell with modulated transcription, where

XX CC the protein is capable of exhibiting at least one of the following

XX CC biological activities, which include enhanced expression of the

XX CC protein in a plant results in a delay in flowering time, introduction of

XX CC the protein into a cell results in modulation of methylation of

XX CC chromosomal DNA in the cell, reduction of expression of the protein in a

XX CC plant results in enhanced endosperm development and expression of the

XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.

XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.

XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates

XX CC transcription of target genes by demethylation. The present

XX CC sequence represents a DMT-like protein.

XX SQ Sequence 1413 AA;

Query Match 63.5%; Score 440; DB 23; Length 1413;

Best Local Similarity 49.7%; Pred. No. 2e-64;

Matches 99; Conservative 0; Mismatches 98; Indels 2; Gaps 2;

Qy 11 EXXXXXXXXXXXXXXXXXXXXGXXXXXXLRXXXXXXFLXXXXXXVXX 70

Db 874 ETRSTMDVDMKAIKRAADVKEVETIKSRGMNHLKLAIRIQLNNKMGFDLRLVND 933

Qy 71 XGIDLEWLRXXXXXXGGLKXGKXECVRLLLXXXXXAFPVDNTNVRGXVXGX 130

Db 934 HGSIDLEWLRDVPDPKAKELLSFNGGLKXSVCEVRLTLHLAFPPVDNTNVRGLGW 993

Qy 131 VPLXPLPXXXXXHLXLYPXXXXXQKXLPRLCKLXQXTLYELHYXXITFGKXXFCTKXX 190

Db 994 VPLQPLPESLQHLLEWYPMLESIQKYLWPRCLKLDQKTLTYELHYQMITFGK-VFCTRSK 1052

Qy 191 PNCNACPMXXECXXXXXA 209

Db 1053 PNCNACPM-KGECRHFASA 1070

RESULT 2

AAU72741

ID AAU72741 standard; Protein; 1114 AA.

XX AC AAU72741;

XX DT 26-FEB-2002 (first entry)

XX DE Arabidopsis DMT4 (IDMT4) protein.

XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;

XX KW DNA demethylation; transgenic plant; transcription modulation;

XX KW flowering time; endosperm development; MEDEA.

XX OS Arabidopsis Thaliana.

XX PN WO200180626-A1.

XX PD 01-NOV-2001.

XX PF 23-APR-2001; 2001WO-US13059.

PR 21-APR-2000; 2000US-0553690.

PA (REGC ) UNIV CALIFORNIA.

XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;

XX DR WPT; 2002-055307/07.

XX DR N-PSDB; AAS96694.

XX PT New polynucleotide that control plant development comprising a sequence

XX PT having a specific homology to DEMETER domains A, B or C

XX PS Disclosure; Page 75; 109pp; English.

XX CC The invention relates to an isolated polynucleotide sequence or their

XX CC complement encoding a polypeptide having a sequence at least 40%

XX CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B

XX CC or C or their combinations. Also included are an expression

XX CC cassette comprising the polynucleotide or comprising a heterologous

XX CC polynucleotide under the control of a promoter at least 70% identical to

XX CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated

XX CC region of DMT, a host cell comprising an exogenous polynucleotide

XX CC encoding a DMT-like protein and a transgenic plant comprising a

XX CC polynucleotide encoding a DMT-like protein. The expression cassette is

XX CC useful for modulating transcription. The method comprises introducing

XX CC the cassette into a host cell preferably Agrobacterium by sexual

XX CC cross, and selecting a host cell with modulated transcription, where

XX CC the protein is capable of exhibiting at least one of the following

XX CC biological activities, which include enhanced expression of the

XX CC protein in a plant results in a delay in flowering time, introduction of

XX CC the protein into a cell results in modulation of methylation of

XX CC chromosomal DNA in the cell, reduction of expression of the protein in a

XX CC plant results in enhanced endosperm development and expression of the

XX CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.

XX CC The polynucleotide is useful for detecting a nucleic acid in a sample.

XX CC DEMETER is related to 5-methylcytosine glycosylases and regulates

XX CC transcription of target genes by demethylation. The present

XX CC sequence represents a DMT-like protein.

XX SQ Sequence 1114 AA;

Query Match 62.5%; Score 433; DB 23; Length 1114;

Best Local Similarity 55.5%; Pred. No. 2.3e-63;

Matches 96; Conservative 0; Mismatches 65; Indels 12; Gaps 3;

Qy 37 IXRGXXXXXXRXRXXXXXXFLXXXXXXGDXIDLEWLRXXXXXXKXLLXXXXG 96

Db 595 IKRGQFRLSERI-----LKFLNDEVNQNGNIDLEWLRNAPSHLVKRYLLEIEG 644

Qy 97 XGLKXSECVRLLLXLLXXAFPPVDNTNVRGXVXVPLXPLPXXXXXQXHLXLYPXXXXXQK 156

Db 645 IGLKSAECVRLGLKHHAFPPVDNTNVRGIAVRGLVPLPLPVGQMHOLFEPSPMSDIQK 704

Qy 157 XLWPRCLKLXQXTLYELHYXXITFGKXXFCTKXXPNACPMXXXXXECXXXXXA 209

Db 705 YLWPRCLKLPQETLYELHYQMITFGK-VFCTKTIPNCNACPM-KSECKYFASA 755

RESULT 3

AAU72736

ID AAU72736 standard; Protein; 1729 AA.

XX AC AAU72736;

XX DT 26-FEB-2002 (first entry)

XX DE Arabidopsis Demeter protein.

XX KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;

XX KW DNA demethylation; transgenic plant; transcription modulation;

XX KW flowering time; endosperm development; MEDEA.

XX OS Arabidopsis Thaliana.

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:08 ; Search time 35.7026 Seconds  
(without alignments)  
858.415 Million cell updates/sec

Title: US-09-840-743-72  
Perfect score: 693  
Sequence: 1 WXXXXXXXEXXXXXXX.....XXXXXXXXXXXXLXXX 230

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	440	63.5	1413	23	Arabidopsis DMT2 (
2	433	62.5	1114	23	Arabidopsis DMT4 (
3	433	62.5	1729	23	Arabidopsis Demete
4	433	62.5	1952	23	Rice DMT1 protein.
5	432	62.3	298	23	Soybean DMT1 prote
6	432	62.3	1332	23	Arabidopsis DMT3 (
7	373.5	53.9	251	23	Soybean DMT2 prote
8	335.5	48.4	1309	23	Herbicidally activ
9	255	36.8	179	23	Soybean DMT protein
10	141	20.3	145	23	Wheat DMT2 protein

11	106	15.3	276	23	AAU72747	Corn DMT3 protein.
12	98	14.1	166	23	AAU72746	Corn DMT2 protein.
13	86	12.4	223	22	AAB96207	Putative P. abyssal
14	74	10.7	218	23	ABB54387	Lactococcus lactis
15	72	10.4	219	23	ABB48015	Listeria monocytog
16	68	9.8	210	22	AAG81939	S. epidermidis ope
17	68	9.8	224	22	AAG82795	S. epidermidis ope
18	67	9.7	224	22	ABP38380	Staphylococcus epi
19	63	9.1	308	22	ABB58227	Drosophila melanog
20	61	8.8	1072	22	ABB69797	Drosophila melanog
21	59	8.5	350	19	AAW60244	Amino acid of the
22	59	8.5	350	22	AAU69755	Escherichia coli (
23	57.5	8.3	118	21	AAB37443	Human cystatin D.
24	57.5	8.3	122	21	AAAY81137	Human wild-type cy
25	57.5	8.3	122	21	AAAY81171	Human mutant cysta
26	57.5	8.3	122	21	AAAY81172	Human mutant cysta
27	57.5	8.3	122	21	AAAY81173	Human mutant cysta
28	57.5	8.3	122	21	AAAY81174	Human mutant cysta
29	57.5	8.3	122	21	AAAY81175	Human mutant cysta
30	57.5	8.3	122	21	AAAY81176	Human mutant cysta
31	57.5	8.3	122	21	AAAY81177	Human mutant cysta
32	57.5	8.3	122	21	AAAY81181	Human mutant cysta
33	57.5	8.3	122	21	AAAY81182	Human mutant cysta
34	57.5	8.3	142	22	AAE11210	Human cystatin D (
35	57.5	8.3	142	22	AAE04437	Human cystatin D h
36	57.5	8.3	142	22	AAE02408	Human cystatin D p
37	57.5	8.3	142	23	AAO15149	Human cystatin D p
38	56.5	8.2	122	21	AAAY81178	Human mutant cysta
39	56	8.1	195	21	AAG12675	Arabidopsis thalia
40	56	8.1	326	21	AAG12674	Arabidopsis thalia
41	55.5	8.0	320	19	AAW70220	Leishmania antigen
42	55.5	8.0	320	19	AAW70236	Leishmania antigen
43	55.5	8.0	320	23	ABG60872	Leishmania antigen
44	55.5	8.0	320	23	ABG60888	Leishmania antigen
45	55.5	8.0	320	23	AAU71816	Leishmania antigen

ALIGNMENTS

RESULT 1  
AAU72737  
ID AAU72737 standard; Protein; 1413 AA.  
XX AC AAU72737;  
XX AC  
DT 26-FEB-2002 (first entry)  
XX DE Arabidopsis DMT2 (IDMT2) protein.  
XX DE  
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX OS Arabidopsis Thaliana.  
XX PN WO200180626-A1.  
XX PD 01-NOV-2001.  
XX PF 23-APR-2001; 2001WO-US13059.  
XX PR 21-APR-2000; 2000US-0553690.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX DR WPI; 2002-055307/07.  
XX DR N-PSDB; AAS96692.  
XX PT New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A,B or C



Db 34 GYLTSSVCVRCVAGKSCPDVANKVCL-GGSEFINNN----- 74  
Qy 155 QKXLWPRCLKQXTLYELHYXXITFGKXFXCTKXXPNCNACPMXXXP 203  
Db 75 -----CPCTDPNCASCPSDAGTC 92

RESULT 14  
US-08-508-550-36  
; Sequence 36, Application US/08808550  
; Patent No. 5871992  
; GENERAL INFORMATION:  
; APPLICANT: Teebor, George W.  
; APPLICANT: Hilbert, Timothy P.  
; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND  
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/808,550  
; FILING DATE: 26-FEB-1997  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1049-1-001 N  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 207 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli

US-08-508-550-36  
Query Match 7.0%; Score 48.5; DB 2; Length 207;  
Best Local Similarity 23.4%; Pred. No. 5.9;  
Matches 22; Conservative 7; Mismatches 38; Indels 27; Gaps 5;

Qy 103 ECVRLLLXXAFVDTNNGRIXVRXGVPLXPLPXXQXHXLXXYPXXXXXQKXLWPR 162  
Db 131 ECV-----GIAVDTHVHRISNRLGWIK-TSTPEKTQKALEILLP-----KSEW--- 172  
Qy 163 CKLXQXTLYELHYXXITFGKXXFXCTKXXPNCNAC 196  
Db 173 -----QPINHLLVGFQMQ-CQVPRPKGTC 197

RESULT 15  
US-08-508-550-39  
; Sequence 39, Application US/08808550  
; Patent No. 5871992  
; GENERAL INFORMATION:

; APPLICANT: Teebor, George W.  
; APPLICANT: Hilbert, Timothy P.  
; TITLE OF INVENTION: MAMMALIAN ENDONUCLEASE III AND  
; TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/808,550  
; FILING DATE: 26-FEB-1997  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1049-1-001 N  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 39:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 259 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: C. elegans

US-08-808-550-39  
Query Match 7.0%; Score 48.5; DB 2; Length 259;  
Best Local Similarity 23.4%; Pred. No. 7.4;  
Matches 22; Conservative 7; Mismatches 38; Indels 27; Gaps 5;  
Qy 103 ECVRLLLXXAFVDTNNGRIXVRXGVPLXPLPXXQXHXLXXYPXXXXXQKXLWPR 162  
Db 131 ECV-----GIAVDTHVHRISNRLGWIK-TSTPEKTQKALEILLP-----KSEW--- 172  
Qy 163 CKLXQXTLYELHYXXITFGKXXFXCTKXXPNCNAC 196  
Db 173 -----QPINHLLVGFQMQ-CQVPRPKGTC 197

Search completed: November 8, 2002, 19:29:16  
Job time : 15.1536 secs

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/183.861  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/022.765  
FILING DATE: 12-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 320 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-183-861-55

Query Match 8.0%; Score 55.5; DB 4; Length 320;  
Best Local Similarity 19.3%; Pred. No. 0.7;  
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;  
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDNVRGXVPLPXXXXXHXHLLXXYPXXXXX 154  
Db 34 GYGLTSSVVCVRCVAGCKSCPVDANVCKVL-GGSEPINNM----- 74  
Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGXXFXCTKXXPNACPMXXEC 203  
Db 75 -----CPCDTPNCASCPSDAGTC 92

RESULT 12  
US-09-022-765-22  
Sequence 22, Application US/09022765  
Patent No. 6375955  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022.765  
FILING DATE: 12-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 320 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-022-765-22  
Query Match 8.0%; Score 55.5; DB 4; Length 320;  
Best Local Similarity 19.3%; Pred. No. 0.7;  
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;  
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDNVRGXVPLPXXXXXHXHLLXXYPXXXXX 154  
Db 34 GYGLTSSVVCVRCVAGCKSCPVDANVCKVL-GGSEPINNM----- 74  
Qy 155 QKXLPRLCKLXQXTLYELHYXXITFGXXFXCTKXXPNACPMXXEC 203  
Db 75 -----CPCDTPNCASCPSDAGTC 92

RESULT 13  
US-09-022-765-55  
Sequence 55, Application US/09022765  
Patent No. 6375955  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, David C.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USEIN THE THERAPY AND  
NUMBER OF SEQUENCES: 87  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022.765  
FILING DATE: 12-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.420C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 320 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-022-765-55

Query Match 8.0%; Score 55.5; DB 4; Length 320;  
Best Local Similarity 19.3%; Pred. No. 0.7;  
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;  
Qy 96 GXGLKSXE-CVRLXLXXXAFPVDNVRGXVPLPXXXXXHXHLLXXYPXXXXX 154



```
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
; FILE REFERENCE: IL-10284
; CURRENT APPLICATION NUMBER: US/09/650,855
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 37
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-650-855-37

Query Match      8.5%; Score 59; DB 4; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.21;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps 3;

QY 96 GXGLKSECVRLXLXXXAPPV-DTNVGRIVRXGVPLPLPXXXXXHXLLXXYPXXXXX 154
Db 116 GVGRSTAGAILSLGRH-FPILDGNVRRVLAR-----CYAVSGWPGKRKEV 160

QY 155 QKXLPRLCKLQXTLVHELHYXITFGKXXFXCTKXXXPNCNACPM 198
Db 161 ENKLWSLEQTPVGVVERFQNMMDLGAMICTRKRKCSLCPL 204

RESULT 6
US-08-744-138-4
; Sequence 4, Application US/08744138
; Patent No. 6011012
; GENERAL INFORMATION:
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ni, Jian
; APPLICANT: Rosen, Craig A.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Human Cystatin E
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,138
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF202P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301 309 8504
; TELEFAX: 301 309 8512
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Cystatin D
US-08-744-138-4
```

```
Query Match      8.3%; Score 57.5; DB 3; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 174 HYXXITFGKXXFCFKXXPCNACP 197
Db 84 YFNVKFGRTT-CTKSQPNLDNCP 106

RESULT 7
US-09-431-480-7
; Sequence 7, Application US/09431480
; Patent No. 6235708
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Feldhaus, Andrew
; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
; FILE REFERENCE: 98-72
; CURRENT APPLICATION NUMBER: US/09/431,480
; PRIOR FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: 60/109,217
; EARLIER FILING DATE: 1998-11-20
; EARLIER APPLICATION NUMBER: 60/156,382
; EARLIER FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-431-480-7

Query Match      8.3%; Score 57.5; DB 4; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 174 HYXXITFGKXXFCFKXXPCNACP 197
Db 84 YFNVKFGRTT-CTKSQPNLDNCP 106

RESULT 8
US-09-617-302-7
; Sequence 7, Application US/09617302
; Patent No. 6245529
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Feldhaus, Andrew
; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
; FILE REFERENCE: 98-72 C1
; CURRENT APPLICATION NUMBER: US/09/617,302
; CURRENT FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/431,480
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: 60/109,217
; PRIOR FILING DATE: 1998-11-20
; PRIOR APPLICATION NUMBER: 60/156,382
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-617-302-7

Query Match      8.3%; Score 57.5; DB 4; Length 142;
Best Local Similarity 41.7%; Pred. No. 0.14;
Matches 10; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY 174 HYXXITFGKXXFCFKXXPCNACP 197
Db 84 YFNVKFGRTT-CTKSQPNLDNCP 197
```

```
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,550
; FILING DATE: 26-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-001 N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus subtilis
; US-08-808-550-31

Query Match          9.2%; Score 64; DB 2; Length 209;
Best Local Similarity 23.1%; Pred. No. 0.02;
Matches 25; Conservative 10; Mismatches 55; Indels 18; Gaps 3;

QY 91 LXXGXGKLSXECVRLKXQTLXLYELHYXXITFGKXXFCTKXXPNACPM 198
Db 112 LVKLFVGVRKTNVVSVAFGVPAIAVDTHVERVSKRLGICRWKDSVLEVEKTLMRKVP- 170
QY 151 XXXXOKLWPRCKLXQTLXLYELHYXXITFGKXXFCTKXXPNACPM 198
Db 171 -----KEDWS-----VTHRLIFFGR-YHCKAQSPRAECPL 201

RESULT 3
US-08-813-574-9
; Sequence 9, Application US/08813574
; Patent No. 6013473
; GENERAL INFORMATION:
; APPLICANT: Wei, Ying-Fei
; TITLE OF INVENTION: Human Muty
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,574
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/013,132
; FILING DATE: 11-MAR-1996
; ATTORNEY/AGENT INFORMATION:

; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-813-574-9

Query Match          8.5%; Score 59; DB 3; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.21;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps 3;

QY 96 GXGLKSXECVRLKXQTLXLYELHYXXITFGKXXFCTKXXPNACPM 198
Db 116 GVGRTAGAILSLSLGKH-FPILDGNVKKVLAR-----CYAVSGWPGKKEV 160
QY 155 QKXLWPRCKLXQTLXLYELHYXXITFGKXXFCTKXXPNACPM 198
Db 161 ENKLWSLSQVTPAVGVVERFNQAMMDLGMICTRSKPKCSLCPL 204

RESULT 4
US-09-651-656-37
; Sequence 37, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
; TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-651-656-37

Query Match          8.5%; Score 59; DB 4; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.21;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps 3;

QY 96 GXGLKSXECVRLKXQTLXLYELHYXXITFGKXXFCTKXXPNACPM 198
Db 116 GVGRTAGAILSLSLGKH-FPILDGNVKKVLAR-----CYAVSGWPGKKEV 160
QY 155 QKXLWPRCKLXQTLXLYELHYXXITFGKXXFCTKXXPNACPM 198
Db 161 ENKLWSLSQVTPAVGVVERFNQAMMDLGMICTRSKPKCSLCPL 204

RESULT 5
US-09-650-855-37
; Sequence 37, Application US/09650855
; Patent No. 6365355
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
```

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	67	9.7	224	4	US-09-134-001C-3225		Sequence 3225, Appl
2	64	9.2	209	2	US-08-808-550-31		Sequence 31, Appl
3	59	8.5	350	3	US-08-813-574-9		Sequence 9, Appl
4	59	8.5	350	4	US-09-651-656-37		Sequence 37, Appl
5	59	8.5	350	4	US-09-650-855-37		Sequence 37, Appl
6	57.5	8.3	142	3	US-08-744-138-4		Sequence 4, Appl
7	57.5	8.3	142	4	US-09-431-480-7		Sequence 7, Appl
8	57.5	8.3	142	4	US-09-617-302-7		Sequence 7, Appl
9	57.5	8.3	142	4	US-09-241-376-4		Sequence 4, Appl
10	55.5	8.0	320	4	US-09-183-861-22		Sequence 22, Appl
11	55.5	8.0	320	4	US-09-183-861-55		Sequence 55, Appl
12	55.5	8.0	320	4	US-09-022-765-22		Sequence 22, Appl
13	55.5	8.0	320	4	US-09-022-765-55		Sequence 55, Appl
14	48.5	7.0	207	2	US-08-808-550-36		Sequence 36, Appl
15	48.5	7.0	207	2	US-08-808-550-39		Sequence 39, Appl
16	47.5	6.9	120	6	5432264-4		Patent No. 5432264
17	47.5	6.9	145	2	US-08-832-535-11		Sequence 11, Appl
18	47.5	6.9	146	2	US-08-791-522-3		Sequence 3, Appl
19	47.5	6.9	146	3	US-08-744-138-3		Sequence 3, Appl
20	47.5	6.9	146	3	US-09-019-485-4		Sequence 4, Appl
21	47.5	6.9	146	3	US-09-314-777-3		Sequence 3, Appl
22	47.5	6.9	146	4	US-09-431-480-6		Sequence 6, Appl
23	47.5	6.9	146	4	US-09-617-302-6		Sequence 6, Appl
24	47.5	6.9	146	4	US-09-241-376-3		Sequence 3, Appl
25	47.5	6.9	146	5	PCT-US95-07135-9		Sequence 9, Appl
26	47.5	6.9	146	6	5432264-6		Patent No. 5432264
27	47	6.8	191	3	US-08-974-022-52		Sequence 52, Appl

**THIS PAGE BLANK (USPTO)**







```
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Calcium Channel Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT013PIC1
; CURRENT APPLICATION NUMBER: US/10/050,786
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 09/774,028
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: PCT/US00/20392
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/145,958
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 60/149,446
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/189,064
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 7
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-050-786-7
```

```
Query Match 7.3%; Score 50.5; DB 9; Length 584;
Best Local Similarity 36.1%; Pred. No. 2.6;
Matches 13; Conservative 2; Mismatches 20; Indels 1; Gaps 1;
```

```
QY 158 LMPRLCKXQXTLYELHYXXITFGKXXFCCKXXPNC 193
      ||| | | | | | | | | | | | | | | |
DB 331 LLPRMASLG-LTLFFYYSFAIVGMEFFCGIVFPNC 365
```

```
RESULT 7
US-09-764-877-1188
; Sequence 1188, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1188
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1188
```

```
Query Match 6.9%; Score 47.5; DB 10; Length 82;
Best Local Similarity 22.2%; Pred. No. 0.91;
Matches 14; Conservative 10; Mismatches 34; Indels 5; Gaps 2;
```

```
QY 116 PVDTNVGRIXRVXGVPXLPLPXXXXXHXLPXXXXXQKXLPRLCKLXQXTLYELHY 175
      | | | | | | | | | | | | | | | | | |
DB 20 PSPHTASRLPIR----PFLVITAYGD-HILRFFPAQDSEDFSLYPMSCVIHIYTLFLVAV 74
```

```
QY 176 XXI 178
;
DB 75 FSV 77
```

```
RESULT 8
US-09-775-932-2
; Sequence 2, Application US/09775932
; Patent No. US20020137671A1
; GENERAL INFORMATION:
; APPLICANT: University of British Columbia
; TITLE OF INVENTION: Production and use of Modified Cystatins
; FILE REFERENCE: 58069
```

```
; CURRENT APPLICATION NUMBER: US/09/775,932
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: CA99/00717
; PRIOR FILING DATE: 1999-08-05
; PRIOR APPLICATION NUMBER: 60/095,503
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-775-932-2
```

```
Query Match 6.9%; Score 47.5; DB 10; Length 120;
Best Local Similarity 33.3%; Pred. No. 1.4;
Matches 8; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
```

```
QY 174 HYXXITFGKXXFCCKXXPNCNACP 197
      :: : | : ||| || : ||
DB 62 YFLDVELGRTT-CTKTQPNLDNCP 84
```

```
RESULT 9
US-09-940-497-3
; Sequence 3, Application US/09940497
; Patent No. US20020052476A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Human Cystatin E
; FILE REFERENCE: PF202FD2
; CURRENT APPLICATION NUMBER: US/09/940,497
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US 09/241,376
; PRIOR FILING DATE: 1999-02-02
; PRIOR APPLICATION NUMBER: US 08/744,138
; PRIOR FILING DATE: 1996-11-05
; PRIOR APPLICATION NUMBER: US 08/461,030
; PRIOR FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-940-497-3
```

```
Query Match 6.9%; Score 47.5; DB 10; Length 146;
Best Local Similarity 33.3%; Pred. No. 1.7;
Matches 8; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
```

```
QY 174 HYXXITFGKXXFCCKXXPNCNACP 197
      :: : | : ||| || : ||
DB 88 YFLDVELGRTT-CTKTQPNLDNCP 110
```

```
RESULT 10
US-09-969-834-3
; Sequence 3, Application US/09969834
; Patent No. US20020102711A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga K.
; TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
```

; PRIOR FILING DATE: 1999-02-02  
; PRIOR APPLICATION NUMBER: US 08/744,138  
; PRIOR FILING DATE: 1996-11-05  
; PRIOR APPLICATION NUMBER: US 08/461,030  
; PRIOR FILING DATE: 1995-06-05  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-940-497-4

Query Match 8.3%; Score 57.5; DB 10; Length 142;  
Best Local Similarity 41.7%; Pred. No. 0.037;  
Matches 10; Conservative 4; Mismatches 1; Gaps 1;

Qy 174 HYXXITFGKXXFCYKXXPNCNACP 197  
Db 84 YFNVKFGRTT-CTKQPNDNCP 106

## RESULT 3

US-09-874-923-22  
; Sequence 22, Application US/09874923  
; Patent No. US20020081320A1

; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; APPLICANT: Brannon, Mark  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; FILE REFERENCE: 210121.420C8  
; CURRENT APPLICATION NUMBER: US/09/874,923  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 320  
; TYPE: PRT  
; ORGANISM: Leishmania major  
US-09-874-923-22

Query Match 8.0%; Score 55.5; DB 10; Length 320;  
Best Local Similarity 19.3%; Pred. No. 0.2;  
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;

Qy 96 GXGLKSXE-CVRLLLXXXAFVDTNVGRXGXVPLPXXXXXHXLLXXYPXXXXX 154  
Db 34 GYGLTSSVCVRCVAGCKSCPVDANVCKVCL-GGSEPINNM----- 74

Qy 155 QKXLWPRCLKXQXTLYELHYXXITFGKXXFCYKXXPNCNACPXXXXXEC 203  
Db 75 -----CPCTDPNCASCPDAGTC 92

## RESULT 4

US-09-874-923-55  
; Sequence 55, Application US/09874923  
; Patent No. US20020081320A1

; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay

; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; APPLICANT: Brannon, Mark  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; FILE REFERENCE: 210121.420C8  
; CURRENT APPLICATION NUMBER: US/09/874,923  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 55  
; LENGTH: 320  
; TYPE: PRT  
; ORGANISM: Leishmania major  
US-09-874-923-55

Query Match 8.0%; Score 55.5; DB 10; Length 320;  
Best Local Similarity 19.3%; Pred. No. 0.2;  
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;

Qy 96 GXGLKSXE-CVRLLLXXXAFVDTNVGRXGXVPLPXXXXXHXLLXXYPXXXXX 154  
Db 34 GYGLTSSVCVRCVAGCKSCPVDANVCKVCL-GGSEPINNM----- 74

Qy 155 QKXLWPRCLKXQXTLYELHYXXITFGKXXFCYKXXPNCNACPXXXXXEC 203  
Db 75 -----CPCTDPNCASCPDAGTC 92

## RESULT 5

US-09-874-923-121  
; Sequence 121, Application US/09874923  
; Patent No. US20020081320A1

; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Webb, John R.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Coler, Rhea  
; APPLICANT: Probst, Peter  
; APPLICANT: Brannon, Mark  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; FILE REFERENCE: 210121.420C8  
; CURRENT APPLICATION NUMBER: US/09/874,923  
; CURRENT FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 121  
; LENGTH: 709  
; TYPE: PRT  
; ORGANISM: Leishmania major and chagasi  
US-09-874-923-121

Query Match 8.0%; Score 55.5; DB 10; Length 709;  
Best Local Similarity 19.3%; Pred. No. 0.47;  
Matches 21; Conservative 7; Mismatches 30; Indels 51; Gaps 3;

Qy 96 GXGLKSXE-CVRLLLXXXAFVDTNVGRXGXVPLPXXXXXHXLLXXYPXXXXX 154  
Db 423 GYGLTSSVCVRCVAGCKSCPVDANVCKVCL-GGSEPINNM----- 463

Qy 155 QKXLWPRCLKXQXTLYELHYXXITFGKXXFCYKXXPNCNACPXXXXXEC 203  
Db 464 -----CPCTDPNCASCPDAGTC 481

## RESULT 6

US-10-050-786-7  
; Sequence 7, Application US/10050786  
; Patent No. US2002015539A1

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:26:04 ; Search time 7.14052 Seconds  
(without alignments)  
464.428 Million cell updates/sec

```

Title: US-09-840-743-72
Perfect score: 693
Sequence: 1 WXXXXXXXEXXXXXDX.....XXXXXXXXXXXXLXXX 230

```

Scoring table: BLOSUM62

Gapex 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- database :   
 1 : /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep : \*   
 2 : /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep : \*   
 3 : /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep : \*   
 4 : /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep : \*   
 5 : /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep : \*   
 6 : /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep : \*   
 7 : /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep : \*   
 8 : /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep : \*   
 9 : /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep : \*   
 10 : /cgn2\_6/ptodata/1/pubaa/US09\_PUBCOMB.pep : \*   
 11 : /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep : \*   
 12 : /cgn2\_6/ptodata/1/pubaa/US10\_PUBCOMB.pep : \*   
 13 : /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep : \*   
 14 : /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	57.5	8.3	122	10	US-09-775-932-10	Sequence 10, Appl
2	57.5	8.3	142	10	US-09-940-497-4	Sequence 4, Appl
3	55.5	8.0	320	10	US-09-874-923-22	Sequence 22, Appl
4	55.5	8.0	320	10	US-09-874-923-55	Sequence 55, Appl
5	55.5	8.0	709	10	US-09-874-923-121	Sequence 121, App
6	50.5	7.3	584	9	US-09-050-786-7	Sequence 7, Appl
7	47.5	6.9	82	10	US-09-764-877-1188	Sequence 1188, Ap
8	47.5	6.9	120	10	US-09-775-932-2	Sequence 2, Appl
9	47.5	6.9	146	10	US-09-940-497-3	Sequence 3, Appl
10	47.5	6.9	145	10	US-09-969-834-3	Sequence 3, Appl
11	46	6.6	486	10	US-09-925-301-1326	Sequence 1326, Ap
12	46	6.6	758	10	US-09-801-368-224	Sequence 224, App
13	45.5	6.6	494	10	US-09-741-669-468	Sequence 468, App
14	45	6.5	84	10	US-09-864-761-35736	Sequence 35736, A
15	45	6.5	121	10	US-09-775-932-6	Sequence 6, Appl
16	45	6.5	141	10	US-09-940-497-7	Sequence 7, Appl
17	44	6.3	106	9	US-09-943-388-9	Sequence 9, Appl
18	44	6.3	106	10	US-09-818-954A-3	Sequence 3, Appl
19	44	6.3	130	9	US-09-943-388-2	Sequence 2, Appl

## ALIGNMENTS

## RESULT 1

US-09-775-932-10  
; Sequence 10, Application US/09775932

Sequence ID: US2002013  
Patent No. US2002013  
GENERAL INFORMATION:

APPLICANT: University of British Columbia

APPLICANT: UNIVERSITY  
; TITLE OF INVENTION: PH  
; FILE REFERENCE: 58069

FILE REFERENCE: 38009  
CURRENT APPLICATION NUMBER: US/09/775,932  
CURRENT FILING DATE: 2001-02-02

; CURRENT FILING DATE: 2001-02-02  
 ; PRIOR APPLICATION NUMBER: CA99/00717  
 ; PRIOR FILING DATE: 1999-08-05

; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: 60/  
; PRIOR FILING DATE: 1999-08-05

; PRIOR FILING DATE: 1998-08-  
 ; NUMBER OF SEQ ID NOS: 32  
 ;

```

; SOFTWARE: P:
;
; SEQ ID NO 10

```

```

; LENGTH: 122
; TYPE: PRT

```

Query Match	8.3%	Score	57.5;	DB	10;	Length	122;
Best Local Similarity	41.7%	Prod	No	0	031.		

Oy 174 HVXXITFGKXXEFTKXXPNACBP 197

QY	I74	HIXXIIFGRXXAF	CIKXXFNCNACF	IS
		:	:	:
		:	:	:
D6	64	YFENYKFCRTT-	CTKSSOBNDNCP	86

## RESULT 2

US-09-940-497-4  
 . Sequence 4 Application US/09940497

; sequence 4, Application US  
; Patent No. US20020052476A1  
: GENERAL INFORMATION.

; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION:

;; TITLE OF INVENTION: Human Cystatin E  
;; FILE REFERENCE: PF202P1D2

;; CURRENT APPLICATION NUMBER: US/0  
;; CURRENT FILING DATE: 2001-08-29

**THIS PAGE BLANK (USPTO)**





```
Query Match      7.4%; Score 51.5; DB 1; Length 312;
Best Local Similarity 40.0%; Pred. No. 1.1;
Matches 14; Conservative 3; Mismatches 17; Indels 1; Gaps 1;
```

```

RESULT 12
YNN2_YEAST
ID YNN2_YEAST STANDARD; PRT; 1056 AA.
AC P53914;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 119.3 kDa protein in FPR1-TOM22 intergenic region.
GN YNL132W OR N1858.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=S288c;
RC MEDLINE=96109932; PubMed=8619318;
RA Mallet L., Bussereau F., Jacquet M.;
RA "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT MP2, CAP/SRV2, NAM9, FKBI/FPR1/RBP1, MOM22 and CPT1, predicts an
RT adenosine deaminase gene and 14 new open reading frames.";
RL Yeast 11:1195-1209(1995).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE UPF0202 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46843; CAA86893.1; -
CC EMBL; Z71408; CAA96014.1; -
CC SGD; S0005076; YNL132W.
DR Hypothetical protein; ATP-binding; Nuclear protein.
KW Hypothetical protein; ATP-binding; Nuclear protein.
FT NP_BIND 283 290 ATP (POTENTIAL).
FT SEQUENCE 1056 AA; 119347 MW; 76721ED0867ED618 CRC64;
Query Match 7.4%; Score 51; DB 1; Length 1056;
Best Local Similarity 30.2%; Pred. No. 4.3;
Matches 13; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

Oy 156 KXLWPRCLKIXQTLXYELHYYXITFTGKXXFCTKXXXPNCACPM 198
| | | | | : | | | | | : | : |
Db 703 KTLPPDLLKLSEOPPHYLHVLGVSYGLTSLHKFKWNKNSFPVP 745

RESULT 13
Y9K_BPP4
ID Y9K_BPP4 STANDARD; PRT; 88 AA.
AC P12552;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Hypothetical 9.7 kDa protein (ORF88) (Putative DNA-binding protein).
OS Bacteriophage P4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC unclassified Myoviridae.
OC NCBI_TaxID=10680;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=90221913; PubMed=2183201;
RA Halling C., Calendar R., Christie G.E., Dale E.C., Deho G., Finkel S.,
RA Flensburg J., Ghisotti D., Kahn M.L., Lane K.B., Lin C.-S.,
RA Lindqvist B.H., Pierson L.S., Six E.W., Sunshine M.G., Ziermann R.;
RT "DNA sequence of satellite bacteriophage P4."
RL Nucleic Acids Res. 18:1649-1649(1990).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=85062840; PubMed=6095206;

```





RESULT 9

MUTY_SALTY	MUTY_SALTY	STANDARD;	PRT;	350 AA.
ID	Q05869;			
AC	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DT	A/G-specific adenine glycosylase (EC 3.2.2.-).			
DE	MUTY OR MUTB OR STM3110.			
GN	Salmonella typhimurium.			
OS	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella.			
OC	NCBI_TaxID=602;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=GW1803;			
RC	MEDLINE=931123173; PubMed=84193300;			
RA	Desiraju V., Shanabrook W.G., Lu A.L.;			
RA	"Nucleotide sequence of the Salmonella typhimurium mutB gene, the			
RT	homolog of Escherichia coli mutY.";			
RL	J. Bacteriol. 175:541-543(1993).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RC	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RL	LT2.";			
RL	Nature 413:852-856(2001).			
CC	-1- FUNCTION: ADENINE GLYCOSYLASE ACTIVE ON G-A AND C-A MISPAIRS.			
CC	-1- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE			
CC	CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER			
CC	POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
EMBL	M86634; AAA27165.1; -.			
DR	EMBL; AE008843; AAL21985.1; -.			
DR	FIR; A40647; A40647.			
DR	HSSP; PI7802; 1MUN.			
DR	StyGene; SGI10238; muty.			
DR	InterPro; IPR004035; EndoIII_FCL.			
DR	InterPro; IPR004036; EndoIII_Hhh.			
DR	InterPro; IPR003265; Endo_3c.			
DR	InterPro; IPR003651; FeS_bind.			
DR	InterPro; IPR003583; HHH_1.			
DR	Pfam; PF00730; Hhh-GPD; 1.			
DR	SMART; SM00478; ENDO3c; 1.			
DR	SMART; SM00525; FES; 1.			
DR	SMART; SM00278; HHH1; 1.			
DR	TIGRFAMs; TIGR01084; muty; 1.			
DR	PROSITE; PS00764; ENONUCLEASE_III_1; 1.			
DR	PROSITE; PS01155; ENONUCLEASE_III_2; 1.			
KW	DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S;			
	Complete proteome.			
FT	METAL 192 192 IRON-SULFUR (4FE-4S) (BY SIMILARITY).			
FT	METAL 199 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).			
FT	METAL 202 202 IRON-SULFUR (4FE-4S) (BY SIMILARITY).			
FT	METAL 208 208 IRON-SULFUR (4FE-4S) (BY SIMILARITY).			
SQ	SEQUENCE 350 AA; 39409 MW; 5C55DBD4D7B7BB69F CRC64;			

Query Match	7.9%	Score 55;	DB 1;	Length 350;	
Best Local Similarity	19.5%	Pred. No. 0.33;			
Matches	22;	Conservative 12;	Mismatches 45;	Indels 34;	Gaps
QY	96	GXGLKXECVRLRLXKXXAFVDFNVGRGXVPLPLPXXQXHXLXKXXVYXXXXXQ	150		
DB	116	GVGRSTAGAILSLALGKHYPILDGNVRVLAR-----CYVSGWPGKKEVE	161		
QY	156	KXILWPRCLKXQXTLYE-----LHYXXITFGKXXFXCTKXXPNACPM	198		
DB	162	NTLW-----TLSQVTPARGVERFNQAMMDLG-AMVCTRSKPKACTLCLP	204		
RESULT 10					
OGGI_HUMAN STANDARD; PRT: 345 AA.					
ID	OGGI_HUMAN	STANDARD;	PRT:	345	AA.
AC	O15527;	O00390;	P78554;	O00670;	O00705;
AC	O9Y6C3;	O9Y6C4;	O9UIK0;	O9UIK1;	O9UIK2;
DT	16-OCT-2001	(Rel. 40,	Created)		
DT	16-OCT-2001	(Rel. 40,	Last sequence update)		
DE	15-JUN-2002	(Rel. 41,	Last annotation update)		
DE	N-glycosylase/DNA	lyase [Includes: 8-oxoguanine	DNA glycosylase		
DE	(EC 3.2.2.-);	DNA-(apurinic or	apryimidinic site) lyase		
DE	(EC 4.2.99.18)	(AP lyase)].			
GN	OGGI OR MMH OR MUTM OR OGGI.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
NCBI_TaxID=9606;					
[1]					
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.					
RC	TISSUE=Colon;				
RA	MEDLINE=97330655;	PubMed=91871114;			
RA	Aburatani H., Hoppo Y., Ishida T., Takashima R., Matsuba C.,				
RA	Kodama T., Takao M., Yasui A., Yamamoto K., Asano M., Fukasawa K.,				
RA	Yoshinari T., Inoue H., Otsuka E., Nishimura S.;				
RT	"Cloning and characterization of mammalian 8-hydroxyguanine-specific				
RT	DNA glycosylase/apurinic, apyrimidinic lyase, a functional mutM				
RT	homologue.";				
RT	Cancer Res. 57:2151-2156(1997).				
RL	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM 1A).				
RP	MEDLINE=97352815;	PubMed=9207108;			
RA	Rosenquist T.A., Zharkov D.O., Grollman A.P.;				
RT	"Cloning and characterization of a mammalian 8-oxoguanine DNA				
RT	glycosylase.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 94:7429-7434(1997).				
RL	[3]				
RP	SEQUENCE FROM N.A. (ISOFORM 2).				
RA	MEDLINE=97368311;	PubMed=9223306;			
RA	Roldan-Arjona T., Wei Y.-F., Carter K.C., Klungland A., Anselmino C.,				
RA	Wang R.-P., Augustus M., Lindahl T.;				
RT	"Molecular cloning and functional expression of a human cDNA encoding				
RT	the antimutator enzyme 8-hydroxyguanine-DNA glycosylase.";				
RT	Proc. Natl. Acad. Sci. U.S.A. 94:8016-8020(1997).				
RL	[4]				
RP	SEQUENCE FROM N.A. (ISOFORM 1A).				
RP	MEDLINE=97368310;	PubMed=9223305;			
RA	Radice J.P., Dherin C., Desmaziere C., Fox M.S., Boiteux S.;				
RT	"Cloning and characterization of hOGG1, a human homolog of the OGG1				
RT	gene of Saccharomyces cerevisiae.";				
RT	Proc. Natl. Acad. Sci. U.S.A. 94:8010-8015(1997).				
RL	[5]				
RP	SEQUENCE FROM N.A. (ISOFORM 1A).				
RP	MEDLINE=97342862;	PubMed=9197244;			
RA	Lu R., Nash H.M., Verdine G.L.;				
RT	"A mammalian DNA repair enzyme that excises oxidatively damaged				
RT	guanines maps to a locus frequently lost in lung cancer.";				
RT	Curr. Biol. 7:397-407(1997).				
RL	[6]				
RP	SEQUENCE FROM N.A. (ISOFORM 1A).				
RP	MEDLINE=97334205;	PubMed=9190902;			



	Matches	17;	Conservative	6;	Mismatches	29;	Indels	1;	Gaps	1;
Qy	76	LEWLRLXXXXXXXXXXLLXGLKSECVRLLLXX-XAFPFVDVTNVGRIXVR	127							
	I:			:	:	:	:	:	:	:
Dd	225	LQQLRVASYEAEHAKLCTLPGVGTVKADCICLMALDKPQAVPDVHHWQIAHR	277							
	I:			:	:	:	:	:	:	:

RESULT 7  
OGGI\_YEAST STANDARD; PRT; 376 AA.

AC P53397;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE N-glycosylase/DNA lyase [Includes: 8-oxoguanine DNA glycosylase  
(EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase  
(EC 4.2.99.18) (AP lyase)].  
GN OGG1 OR YML060W OR YM9958.02.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FL100;  
RX MEDLINE=96224259; PubMed=8643552;  
RA van der Kemp P.A., Thomas D., Barbey R., de Oliveira R., Boiteux S.;  
RT "Cloning and expression in Escherichia coli of the OGG1 gene of  
Saccharomyces cerevisiae, which codes for a DNA glycosylase that  
excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine.";  
RT Proc. Natl. Acad. Sci. U.S.A. 93:5197-5202(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Devlin K., Churcher C., Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP CHARACTERIZATION, AND MUTAGENESIS OF LYS-241.  
RX MEDLINE=97398538; PubMed=9241232;  
RA Girard P.M., Guibourt N., Boiteux S.;  
RT "The Ogg1 protein of Saccharomyces cerevisiae: a 7,8-dihydro-8-oxoguanine DNA glycosylase/AP lyase whose lysine 241 is a critical residue for catalytic activity.";  
RT Nucleic Acids Res. 25:3204-3211(1997).  
RN [4]  
RP CHARACTERIZATION, AND MUTAGENESIS OF LYS-241.  
RX MEDLINE=20143487; PubMed=10677220;  
RA Guibourt N., Castaing B., Van Der Kemp P.A., Boiteux S.;  
RT "Catalytic and DNA binding properties of the ogg1 protein of Saccharomyces cerevisiae: comparison between the wild type and the K241R and K241Q active-site mutant proteins.";  
RT Biochemistry 39:1716-1724(2000).  
CC -I- FUNCTION: DNA REPAIR ENZYME THAT INCISES DNA AT 8-OXOG RESIDUES. EXCESSES 7,8-DIHYDRO-8-OXOGUANINE AND 2,6-DIAMINO-4-HYDROXY-5-N-METHYLFORMAMIDOPYRIMIDINE (FAPY) FROM DAMAGED DNA. HAS A BETA-LYASE ACTIVITY THAT NICKS DNA 3' TO THE LESION.  
CC -I- CATALYTIC ACTIVITY: Endonucleolytic cleavage near apurinic or apyrimidinic sites to products with 5'-phosphate.  
CC -I- SUBCELLULAR LOCATION: Nuclear.  
CC -I- SIMILARITY: BELONGS TO THE OGG1 FAMILY 1.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; U44855; AAC49312.1; -.  
DR EMBL; Z46729; CAAB6715.1; -.  
SGD; S0004525; OGG1.



```

DR ENBL; U28377; AAC69128.1; -
DR ENBL; AE000378; AAC75998.1; -
DR PIR; JQ0546; JQ0546.
DR PIR; B38535; B38535.
DR PDB; 1MUD; 29-OCT-99.
DR PDB; 1MUN; 26-AUG-99.
DR PDB; 1MUY; 20-AUG-99.
DR ECoGene; EG10627; muty.
DR InterPro; IPR004035; EndoIII_FCL.
DR InterPro; IPR004036; EndoIII_HHH.
DR InterPro; IPR003265; Endo_3c.
DR InterPro; IPR003651; FeS_bind.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00730; HhH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR SMART; SM00525; PES; 1.
DR SMART; SM00278; HHH1; 1.
DR TIGRfam; TIGR01084; muty; 1.
DR PROSITE; PS00764; ENDONUCLEASE_III_1; 1.
DR PROSITE; PS01155; ENDONUCLEASE_III_2; 1.
DR DNA repair; Hydrolase; Glycosidase; Iron-sulfur; 4Fe-4S; 3D-structure
KW Complete proteome.
FT METAL 192 192 IRON-SULFUR (4FE-4S).
FT METAL 199 199 IRON-SULFUR (4FE-4S).
FT METAL 202 202 IRON-SULFUR (4FE-4S).
FT METAL 208 208 IRON-SULFUR (4FE-4S).
SQ SEQUENCE 350 AA; 39149 MW; C7D3657C03EBBF4F CRC64;

Query Match 8.5%; Score 59; DB 1; Length 350;
Best Local Similarity 20.2%; Pred. No. 0.071;
Matches 21; Conservative 14; Mismatches 53; Indels 16; Gaps

QY 96 GXLGKXECVRLRLXAXAPPV-DTNVGRIXVRXGXVPLXPLPXXXXXHXLXVYXXXXX 154
DB 116 GVRSTAGAILSLSLGKH-FPLDGNVXKVLAR-----CYAVSGWFGKKVEV 160

QY 155 QKXLPRLCLXÖTYLHYLXXITFGKXXXCTKXXPCNACPM 198
DB 161 ENKLWSLSEQVTPAVGVRFNQAMMDLGAMICTRSGKPKSLCPL 204

RESULT 4
CYTD_HUMAN STANDARD; PRT; 142 AA.
ID CYTD_HUMAN STANDARD; PRT; 142 AA.
AC P28325;
DC 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DT Cystatin D precursor.
GN CST5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041895; PubMed=1939105;
RA Freije J.P., Abrahamson M., Olafsson I., Velasco G., Grubb A.,
RA Lopez-Otin C.;
RT "Structure and expression of the gene encoding cystatin D, a novel
RT human cysteine proteinase inhibitor."
RL J. Biol. Chem. 266:20538-20543(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93340179; PubMed=8340398;
RA Freije J.P., Balbin M., Abrahamson M., Velasco G., Dalboge H.,
RA Grubb A., Lopez-Otin C.;
RT "Human cystatin D. cDNA cloning, characterization of the Escherichia
RT coli expressed inhibitor, and identification of the native protein in
RT saliva."
RL J. Biol. Chem. 268:15737-15744(1993).
RN [3]

```

RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitnegger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis";  
RL Nature 390:249-256(1997).  
CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,  
CC leaving a 3'-terminal unsaturated sugar and a product with a  
CC terminal 5'-phosphate.  
CC -!- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE  
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER  
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U11289; AAA80005.1; -;  
DR EMBL; L47709; AAB38457.1; -;  
DR EMBL; Z99115; CAB14150.1; -;  
DR HSSP; P20625; 2ABK.  
DR Subtilist; BG10956; nth.  
DR InterPro; IPR004035; EndoIII\_FCL.  
DR InterPro; IPR004036; EndoIII\_Hhh.  
DR InterPro; IPR003265; Endo\_3c.  
DR InterPro; IPR003651; Fes\_bind.  
DR InterPro; IPR003583; HHH\_1.  
DR Pfam; PF00730; Hhh-GPD; 1.  
DR SMART; SM00478; Endo3c; 1.  
DR SMART; SM00525; FES; 1.  
DR SMART; SM00278; Hhh1; 1.  
DR TIGRfams; TIGR01083; nth; 1.  
DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; 1.  
DR PROSITE; PS01155; ENDONUCLEASE\_III\_2; 1.  
KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
FT METAL 189  
FT METAL 196 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 199 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 205 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 219 AA; 25000 MW; C38C66DA7948BA40 CRC64;  
  
Query Match 9.2%; Score 64; DB 1; Length 219;  
Best Local Similarity 23.1%; Pred. No. 0.0068;  
Matches 25; Conservative 10; Mismatches 55; Indels 18; Gaps 3;  
  
Qy 91 LLXXXGXLKSKXECVRLRLXLLXXXAFPPDVTNVGRITXVRGXVPLPLPXXXXXHLXXYPX 150  
Db 112 LVKLPVGGRKTANVVSVAFGPAIADVTHVERVSKRIGICRWKDSVLEVEKTLMRKVP- 170  
Qy 151 XXXXQKXLPRLCKLXQXTLYELHXXITFGKXXFXCTKXXPCNACPM 198  
Db 171 -----KEDWS-----VTHRLIFFGR-YHCKAQSPRCAECPL 201  
  
RESULT 2  
END3\_RICPR  
ID\_END3\_RICPR STANDARD; PRT; 212 AA.  
AC 005956;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site)  
DE lyase).  
GN NTH OR RP746.

OS Rickettsia prowazekii.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=782;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RC MEDLINE=97419517; PubMed=9274032;  
RA Andersson J.O.; Andersson S.G.E.;  
RT "Genomic rearrangements during evolution of the obligate  
RT intracellular parasite Rickettsia prowazekii as inferred from an  
RT analysis of 52015 bp nucleotide sequence";  
RL Microbiology 143:2783-2795(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Madrid E;  
RC MEDLINE=99039499; PubMed=9823893;  
RA Andersson S.G.E.; Zomorodipour A.; Andersson J.O.;  
RA Sierheit-Ponten T.; Alsmark U.C.M.; Podowski R.M.; Naeslund A.K.;  
RA Eriksson A.-S.; Winkler H.H.; Kurland C.G.;  
RT "The genome sequence of Rickettsia prowazekii and the origin of  
RT mitochondria";  
RL Nature 396:133-140(1998).  
CC -!- FUNCTION: HAS BOTH AN APURINIC AND/OR APYRIMIDINIC ENDONUCLEASE  
CC ACTIVITY AND A DNA N-GLYCOSYLASE ACTIVITY. INCISES DAMAGED DNA AT  
CC CYTOSINES, THYMINE AND GUANINES. ACTS ON A DAMAGED STRAND, 5'  
CC FROM THE DAMAGED SITE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or  
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,  
CC leaving a 3'-terminal unsaturated sugar and a product with a  
CC terminal 5'-phosphate.  
CC -!- COFACTOR: BINDS A 4FE-4S CLUSTER WHICH IS NOT IMPORTANT FOR THE  
CC CATALYTIC ACTIVITY, BUT WHICH IS PROBABLY INVOLVED IN THE PROPER  
CC POSITIONING OF THE ENZYME ALONG THE DNA STRAND (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE NTH/MUTY FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y11778; CAA72458.1; -;  
DR EMBL; AJ235273; CAA15174.1; -;  
DR HSSP; P20625; 2ABK.  
DR InterPro; IPR004035; EndoIII\_FCL.  
DR InterPro; IPR004036; EndoIII\_Hhh.  
DR InterPro; IPR003265; Endo\_3c.  
DR InterPro; IPR003651; Fes\_bind.  
DR Pfam; PF00730; Hhh-GPD; 1.  
DR SMART; SM00478; Endo3c; 1.  
DR SMART; SM00525; FES; 1.  
DR SMART; SM00278; Hhh1; 1.  
DR TIGRfams; TIGR01083; nth; 1.  
DR PROSITE; PS00764; ENDONUCLEASE\_III\_1; 1.  
DR PROSITE; PS01155; ENDONUCLEASE\_III\_2; 1.  
KW Hydrolase; Nuclease; Endonuclease; Multifunctional enzyme; DNA repair;  
KW Glycosidase; Lyase; Iron-sulfur; 4Fe-4S; Complete proteome.  
FT METAL 187  
FT METAL 187 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 194 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 197 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
FT METAL 203 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
SQ SEQUENCE 212 AA; 24174 MW; A7FDA4F6033A6382 CRC64;  
  
Query Match 9.2%; Score 63.5; DB 1; Length 212;  
Best Local Similarity 24.1%; Pred. No. 0.0079;  
Matches 26; Conservative 10; Mismatches 53; Indels 19; Gaps 3;  
  
Qy 91 LLXXXGXLKSKXECVRLRLXLLXXXAFPPDVTNVGRITXVRGXVPLPLPXXXXXHLXXYPX 150  
Db 112 LVKLPVGGRKTANVVSVAFGPAIADVTHVERVSKRIGICRWKDSVLEVEKTLMRKVP- 170  
Qy 151 XXXXQKXLPRLCKLXQXTLYELHXXITFGKXXFXCTKXXPCNACPM 198  
Db 171 -----KEDWS-----VTHRLIFFGR-YHCKAQSPRCAECPL 201  
  
RESULT 2  
END3\_RICPR  
ID\_END3\_RICPR STANDARD; PRT; 212 AA.  
AC 005956;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site)  
DE lyase).  
GN NTH OR RP746.

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	64	9.2	219	1	END3_BACSU	P03788 bacillus su
2	63.5	9.2	212	1	END3_RICPR	O05956 rickettsia
3	59	8.5	350	1	MUTY_ECOLI	P17802 escherichia
4	57.5	8.3	142	1	CYTD_HUMAN	P28325 homo sapien
5	56.5	8.2	213	1	END3_THEMA	Q9wyK0 thermotoga
6	56.5	8.2	345	1	OGG1_RAT	O70249 rattus norv
7	56.5	8.2	376	1	OGG1_YEAST	P53397 saccharomyc
8	55.5	8.0	345	1	OGG1_MOUSE	O08760 mus musculu
9	55	7.9	350	1	MUTY_SALTY	O05869 salmonella
10	53.5	7.7	345	1	OGG1_HUMAN	O15527 homo sapien
11	51.5	7.4	312	1	OGG1_METHY	O27397 methanobact
12	51	7.4	1056	1	YNN2_YEAST	P53914 saccharomyc
13	49.5	7.1	88	1	Y9K_BP4	P12552 bacterioph
14	49	7.1	1033	1	YDK9_SCHPO	P87115 schizosacch
15	48.5	7.0	259	1	END3_CAEEL	P54137 caenorhabdi
16	47.5	6.9	146	1	CYTC_HUMAN	P01034 homo sapien
17	47.5	6.9	146	1	CYTC_MACMU	O19092 macaca mula
18	47	6.8	256	1	TNR9_MOUSE	P20334 mus musculu
19	47	6.8	1379	1	MK3_MOUSE	O35099 mus musculu
20	46.5	6.7	146	1	CYTC_SAISC	O19093 saimiri sci
21	46.5	6.7	211	1	END3_TREPA	O83754 treponema p
22	46.5	6.7	333	1	ETFA_HUMAN	P13804 homo sapien
23	46.5	6.7	333	1	ETFA_MOUSE	Q991C5 mus musculu
24	46	6.6	424	1	CUTL_CAEEL	O03755 caenorhabdi
25	46	6.6	461	1	MYH_SCHPO	Q10159 schizosacch
26	46	6.6	603	1	YA07_SCHPO	Q10086 schizosacch
27	46	6.6	758	1	YK38_HUMAN	O03825 saccharomyc
28	46	6.6	4488	1	RPA_CVMCH	P19751 murine coro
29	45.5	6.6	113	1	GTHB_MURCI	P12837 muraesox
30	45.5	6.6	140	1	GTH2_ANGAN	P27767 anguilla na
31	45.5	6.6	221	1	GTH2_METTF	P29588 methanobact
32	45.5	6.6	360	1	MUTY_BUCAI	P57617 buchnera ap
33	45.5	6.6	350	1	VP3_ARMV	P24820 rubra mosa

## ALIGNMENTS

34	45.5	6.6	457	1	YB02_ECOLI	esch
35	45	6.5	141	1	CTYN_HUMAN	homo
36	44.5	6.4	249	1	YB4E_SCHPO	schizo
37	44.5	6.4	327	1	OGG1_DROME	drosophi
38	44.5	6.4	438	1	GLGA_THECA	thermus
39	44.5	6.4	2052	1	FVY1_MOUSE	mus muscu
40	44	6.3	151	1	PA2X_RAT	rattus norv
41	44	6.3	275	1	VA16_VACCV	vaccinia vi
42	44	6.3	344	1	YG13_METJA	methanoco
43	44	6.3	377	1	VA16_VARV	variola vir
44	44	6.3	378	1	VA16_VACCC	vaccinia vi
45	44	6.3	486	1	GATB_AERPE	aeropyrum p



**THIS PAGE BLANK (USPTO)**

FT REPEAT 2374 2472 CADHERIN 22.  
FT REPEAT 2481 2576 CADHERIN 23.  
FT REPEAT 2586 2686 CADHERIN 24.  
FT REPEAT 2695 2793 CADHERIN 25.  
FT REPEAT 2804 2897 CADHERIN 26.  
FT REPEAT 2927 3020 CADHERIN 27.  
FT VARIANT 1131 1153 MISSING (IN ALLELE D1).  
FT CONFLICT 1070 1070 I -> V (IN REF. 3).  
FT CONFLICT 1490 1490 S -> R (IN REF. 3).  
FT CONFLICT 1636 1636 S -> G (IN REF. 3).  
FT CONFLICT 1692 1692 P -> S (IN REF. 3).  
FT CONFLICT 1804 1804 I -> V (IN REF. 3).  
FT CONFLICT 2029 2029 I -> L (IN REF. 3).  
FT CONFLICT 2210 2210 A -> P (IN REF. 3).  
FT CONFLICT 2289 2289 S -> A (IN REF. 3).  
FT CONFLICT 2536 2536 T -> S (IN REF. 3).  
FT CONFLICT 2862 2862 Q -> R (IN REF. 3).  
FT CONFLICT 3038 3038 G -> S (IN REF. 3).  
SQ SEQUENCE 3503 AA; 379721 MW; F256924D5FB5F7AE CRC64;

Query Match 6.3%; Score 37; DB 5; Length 3503;  
Best Local Similarity 33.3%; Pred. NO. 1.4e+02;  
Matches 7; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 120 DXHXLXXDXXXYLXIXWP 140  
DB 388 DGHFALTTRDNSIYLVHLP 408

Search completed: November 8, 2002, 19:28:42  
Job time : 40.2614 secs

RA RP SEQUENCE FROM N.A.  
 RA Lennard N.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology";  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z75552; CAA99942.1; -;  
 DR InterPro: IPR000822; Znf\_C2H2.  
 DR Pfam: PF00096; zf-C2H2; 1.  
 DR SMART: SM00355; Znf\_C2H2; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW DNA-binding; Zinc-finger  
 SQ SEQUENCE 624 AA; 71971 MW; 39E9BD2838A9C8D3 CRC64;  
  
 Query Match 6.3%; Score 37; DB 5; Length 624;  
 Best Local Similarity 64.3%; Pred. No. 23;  
 Matches 9; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
  
 QY 109 LRTEHYXXLPDXH 122  
 || || | || |  
 Db 90 LRIEHDV--LPDVH 101  
  
 RESULT 15  
 Q24292  
 ID Q24292 PRELIMINARY; PRT; 3503 AA.  
 AC Q24292; Q9VPS4;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE DACHSOUS protein precursor (ADHERIN).  
 GN DACHSOUS  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=EMBRYO;  
 RX MEDLINE=95324813; PubMed=7601355;  
 RA Clark H.F., Brentrup D., Schneitz K., Bieber A., Goodman C., Noll M.;  
 RT "Dachsous encodes a member of the cadherin superfamily that controls  
 RT imaginal disc morphogenesis in Drosophila.";  
 RL Genes Dev. 9:1530-1542(1995).  
 RN [2]  
 RP REVISIONS.  
 RA Noll M.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of Drosophila melanogaster";  
 Science 287:2185-2195(2000).  
 RL  
 CC -1- FUNCTION: INVOLVED IN MORPHOGENESIS. MAY ALSO BE INVOLVED IN CELL  
 CC ADHESION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN EMBRYONIC ECTODERM. IN LARVAE,  
 CC EXPRESSION IS RESTRICTED TO IMAGINAL DISKS AND BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYOGENESIS WHERE IT  
 CC IS FIRST DETECTED DURING GASTRULATION. ALSO EXPRESSED IN LARVAE  
 CC AND ADULTS.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 DR EMBL: L08811; AAA79129.2; -;  
 DR EMBL: AE003588; AAF51468.2; -;  
 DR HSPD; P15116; INCGJ.  
 DR FlyBase; FBgn0000497; ds.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR00233; Cadherin\_C-term.  
 DR Pfam; PF00028; cadherin; 26.  
 DR Pfam; PF01049; Cadherin\_C-term; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR SMART; SM00112; CA; 26.  
 DR PROSITE; PS00232; CADHERIN\_1; 19.  
 DR PROSITE; PS0268; CADHERIN\_2; 27.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
 KW Polymorphism; Signal  
 FT SIGNAL 1  
 FT CHAIN 21 3503  
 FT DOMAIN 21 3045  
 FT TRANSMEM 3046 3066  
 FT DOMAIN 3067 3503  
 FT REPEAT 58 121  
 FT REPEAT 130 233  
 FT REPEAT 242 340  
 FT REPEAT 353 451  
 FT REPEAT 460 558  
 FT REPEAT 567 662  
 FT REPEAT 671 774  
 FT REPEAT 783 878  
 FT REPEAT 887 983  
 FT REPEAT 992 1100  
 FT REPEAT 1109 1203  
 FT REPEAT 1213 1311  
 FT REPEAT 1320 1431  
 FT REPEAT 1440 1548  
 FT REPEAT 1563 1665  
 FT REPEAT 1674 1793  
 FT REPEAT 1803 1899  
 FT REPEAT 1907 2003  
 FT REPEAT 2012 2146  
 FT REPEAT 2122 2269  
 FT REPEAT 2261 2357

```

OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
[1]
RP SEQUENCE FROM N.A.
RA Murphy L., Harris D.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL035480; CAB36627.1; -.
DR EMBL; AL583925; CAC31872.1; -.
DR Leproma; ML2356; -.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt_C; 1.
DR Pfam; PF00550; pp-binding; 1.
DR PROSITE; PS50075; ACP_DOMAIN; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR Phosphopantetheine; Transferase; Complete proteome.
KW Phosphopantetheine; Transferase; Complete proteome.
SQ SEQUENCE 1540 AA; 164209 MW; D7042CAAE72375 CRC64;

Query Match 6.4%; Score 38; DB 16; Length 1540;
Best Local Similarity 36.4%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

Oy ~ 112 EHVXXLPDXHXXLXXXXXX 133
:| | | | | | | | | |
Db 990 DHDVSTLHDLNALRGVDNVLY 1011

RESULT 12
Q92SK4
ID Q92SK4 PRELIMINARY; PRT; 214 AA.
AC Q92SK4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein R00374.
GN R00374 OR SMC01146.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gouzie T., Goffeau A., Kahn D., Kiss E., Leilaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591783; CAC41811.1; -.

```

```

DR InterPro; IPR002637; Hamlp_like.
DR Pfam; PF01725; Hamlp_like; 1.
DR TIGRFAMS; TIGR00042; Hamlp_like; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 23109 MW; A282C7539CBBCDD CRC64;

Query Match 6.3%; Score 37; DB 16; Length 214;
Best Local Similarity 31.4%; Pred. No. 7.1;
Matches 11; Conservative 3; Mismatches 15; Indels 6; Gaps 2;

Oy 110 RTEHXVXXL----PDHXXLXXXXXXLLXIWP 140
:| | | | | | | | | |
Db 125 RTARFVSVLCCLAWPDGHVELRGEVEGYV--VWPP 157

RESULT 13
Q966M0
ID Q966M0 PRELIMINARY; PRT; 325 AA.
AC Q966M0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein C55B7.11.
GN C55B7.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Wamsley P.;
RT "The sequence of C. elegans cosmid C55B7.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006625; AAK68271.1; -.
SQ SEQUENCE 325 AA; 37141 MW; 55EAA4186BC645C0 CRC64;

Query Match 6.3%; Score 37; DB 5; Length 325;
Best Local Similarity 36.0%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

Oy 109 LRTEHXVXXLPDXHXXLXXXXXX 133
:| | | | | | | | | |
Db 140 LNTETLIEILPDGKILNPRDASKY 164

RESULT 14
Q23159
ID Q23159 PRELIMINARY; PRT; 624 AA.
AC Q23159;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE W04D2.4 protein.
GN W04D2.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Adayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,  
RA Beeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003519; AAF49220.1; -.  
DR FlyBase: FBgn0014075; Ugt.  
DR InterPro: IPR002495; GT\_8.  
DR Pfam: PF01501; Glyco\_transf.8; 1.  
SQ SEQUENCE 1548 AA; 174352 MW; 87D782E7B27E0BE9 CRC64;

Query Match 6.9%; Score 41; DB 5; Length 1548;  
Best Local Similarity 28.2%; Pred. No. 7.3;  
Matches 11; Conservative 2; Mismatches 26; Indels 0; Gaps 0;

QY 96 PXXXXXXXKXKXTEHXXVXLPDXHXXLXXXXXXYL 134  
DB 935 PQTQTRFKLPTDLKTDHSVVKLPKQENLPHFDVAVL 973

RESULT 9  
Q93CK3 PRELIMINARY; PRT; 242 AA.  
AC Q93CK3;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Putative spectinomycin adenylyltransferase.  
OS Enterococcus faecalis (Streptococcus faecalis).  
OC Plasmid pEF418.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Enterococcaceae; Enterococcus.  
OX NCBI\_TaxID=1351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=418;  
RA Gentry-Weeks C.R., Keith J.M., Estay M.T., Pikis A.;  
RT "A resident plasmid of *Enterococcus faecalis* 418 contains multiple  
RT IS1216v elements, aminoglycoside and lincosamide resistance genes, and  
RT a putative ABC transporter gene with similarity to the *Streptococcus*  
RT pneumoniae macrolide efflux pump gene.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF408195; AAL05551.1; -.  
DR InterPro: IPR002934; NTP\_transf.  
DR Pfam: PF01909; NTP\_transf\_2; 1.  
KW Plasmid; transferase.  
SQ SEQUENCE 242 AA; 27360 MW; 503A1E863E48B90C CRC64;

Query Match 6.4%; Score 38; DB 2; Length 242;  
Best Local Similarity 34.6%; Pred. No. 4.8;  
Matches 9; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

QY 109 LRTEHXXVXLPDXHXXLXXXXXXYL 134  
DB 179 LAEWAIPQLPDEHATLLNKAREAYL 204

RESULT 10  
Q49932 PRELIMINARY; PRT; 1446 AA.  
ID Q49932;  
AC Q49932;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Putative polyketide synthase PKSC (PKS).  
GN PKSC OR L518\_F1\_2.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Robison K., Smith D.R.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE  
CC SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN  
CC SECONDARY METABOLISM.  
CC -!- COFACTOR: CONTAINS THREE COVALENTLY BOUND PHOSPHOPANTHETHEINES  
CC (POTENTIAL).  
CC -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).  
CC -!- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT  
CC BINDING OF AMP TO THEIR SUBSTRATE.  
EMBL: U00023; AAA17356.1; -.  
DR InterPro: IPR001227; Ac\_transferase.  
DR InterPro: IPR000794; Ketoacyl-synt.  
DR Pfam: PF00698; Acyl\_transf; 1.  
DR Pfam: PF00109; ketoacyl-synt; 1.  
DR Pfam: PF02801; ketoacyl-synt\_C; 1.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 1.  
KW Hypothetical protein; Transferase; Acyltransferase;  
KW Antibiotic biosynthesis; NADP; Phosphopantetheine;  
KW Multifunctional enzyme; Repeat; Ligase.  
FT ACT\_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).  
FT ACT\_SITE 651 651 ACYL TRANSFERASES (BY SIMILARITY).  
SQ SEQUENCE 1446 AA; 153871 MW; 5D23A2E5E6155F4E CRC64;

Query Match 6.4%; Score 38; DB 2; Length 1446;  
Best Local Similarity 36.4%; Pred. No. 33;  
Matches 8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 112 EHXXVXLPDXHXXLXXXXXXYL 133  
DB 990 DHVSTLHDLHNLRGVDNVL 1011

RESULT 11  
Q9S384 PRELIMINARY; PRT; 1540 AA.  
ID Q9S384;  
AC Q9S384;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Putative polyketide synthase.  
GN ML2356 OR MLCB12.01C.

[illegible]

```
Db 95 -GGKASGKMCDETCSECSIREANSQTVRGTTLLIPORTAMRGSPPLNGTVFQVNEFPAD 153
Qy 218 HXXXXXXPIXXXXXXLXRXXXXXXGXXXXXXLXRXXXXXXGXXXXXCRFXFXXXXX 277
    || || || || || || || || || || || || || || || || || || || || ||
Db 154 HESSLKPIDVPRDWLWLDLPRRTVVFGTSTVTSIFRGLSTEQIQCFWKFGVCVGRFGEQKTR 213
Qy 278 XPXXLXXLHXXXXXK 292
    | | | | |
Db 214 APRPLMARLHFPASK 228

RESULT 2
Q9FTQ2 PRELIMINARY; PRT; 274 AA.
AC Q9FTQ2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0665D10.14 protein.
GN P0665D10.14.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC- Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0665D10.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002861; BAB16489.1; -.
SQ SEQUENCE 274 AA; 30951 MW; D7C826BD3982592D CRC64;

Query Match 24.6%; Score 145; DB 10; Length 274;
Best Local Similarity 26.2%; Pred. No. 2e-24;
Matches 51; Conservative 0; Mismatches 134; Indels 10; Gaps 3;

Qy 104 KXXXXLRTHEHXXVXXLPDXHXXL-----XXDXXXXYLXIXWXPXXXXXXHXXHXXCX 157
    || || || || || || || || || || || || || || || || || || || || ||
Db 67 KNVSLRTEHQVYELPDHPLLEGFNQRPDDPCPYLLSIWTP-GETAQSDTAPKXVC-- 123
Qy 158 XXXXXXXXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCX 217
    || || || || || || || || || || || || || || || || || || || || ||
Db 124 -NSQENGELCASNTCFSCNSIREAQAKVRGTTLLIPCRTAMRGSPPLNGTVFQVNEFPAD 182
Qy 218 HXXXXXXPIXXXXXXLXRXXXXXXGXXXXXXLXRXXXXXXGXXXXXCRFXFXXXXX 277
    || || || || || || || || || || || || || || || || || || || || ||
Db 183 HDSSRNPIDVPRSWIWLPRRTVYFGTSTVTSIFRGLTTEIQQHCFWRFVGVGRFDRTSR 242
Qy 278 XPXXLXXLHXXXXXK 292
    | | | | |
Db 243 APRPLMARLHFPASK 257

RESULT 3
Q9SR66 PRELIMINARY; PRT; 1309 AA.
AC Q9SR66;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T22K18.18 protein.
GN T22K18.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;

Qy 104 KXXXXLRTHEHXXVXXLPDXHXXLX-----XXDXXXXYLXIXWXPXXXXXXHXXHXXCX 157
    || || || || || || || || || || || || || || || || || || || || ||
Db 1109 KIKEKLRTHEHVFELPDHHSILGFERREAEIDVPLLAIWTP-GETVNSIQPPKQRC-- 1165
Qy 158 XXXXXXXXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCXCCXXCX 217
    || || || || || || || || || || || || || || || || || || || || ||
Db 1166 ALFESNNTLCNENKFCQCNKTRREESQTVRGTTLLIPCRTAMRGSPPLNGTVFQVNEFPAD 1225
Qy 218 HXXXXXXPIXXXXXXLXRXXXXXXGXXXXXXLXRXXXXXXGXXXXXCRFXFXXXXX 265
    || || || || || || || || || || || || || || || || || || || || ||
Db 1226 HDSSINPIDVPTLWDLKRRVAVLGSSVSSICKLSVEAIKYNFQEG 1273

RESULT 4
Q9SJQ6 PRELIMINARY; PRT; 1207 AA.
AC Q9SJQ6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ATg36490 protein.
GN ATg36490.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL thaliana.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006919; AAD24633.1; -.
DR InterPro; IPR003651; Fes_bind.
DR SMART; SM00525; FES; 1.
SQ SEQUENCE 1207 AA; 135916 MW; ADDB6C9EE94DDF67 CRC64;

Query Match 15.4%; Score 91; DB 10; Length 1207;
Best Local Similarity 29.8%; Pred. No. 2.1e-11;
Matches 28; Conservative 0; Mismatches 56; Indels 10; Gaps 3;

Qy 104 KXXXXLRTHEHXXVXXLPDXHXXLX-----XXDXXXXYLXIXWXPXXXXXXHXXHXXCX 157
    || || || || || || || || || || || || || || || || || || || || ||
```





**THIS PAGE BLANK (USPTO)**

XX Example 2; SEQ ID NO 6504; 10078pp; English.  
PS  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 279 AA;  
Query Match 15.3%; Score 42; DB 22; Length 279;  
Best Local Similarity 41.2%; Pred. No. 32;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 49 QGXRFXXXWKGSVVDSV 65  
Db 81 EGDPTQWKGVLDQV 97

Search completed: November 8, 2002, 19:25:54  
Job time : 16.9706 secs

CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX SQ Sequence 232 AA;  
 Query Match 15.3%; Score 42; DB 22; Length 232;  
 Best Local Similarity 41.2%; Pred. No. 25;  
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 49 QGXRFXXXWKGVSVDV 65  
 : | | | | | | | | | |  
 Db 34 EGDEPITQWKGTVDQV 50

RESULT 14  
 AAM39788  
 ID -AAM39788 standard; Protein: 258 AA.

XX AC AAM39788;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polypeptide SEQ ID NO 2933.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX OS Homo sapiens.  
 XX PN WO200153312-A1.  
 XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.  
 XX PR 25-APR-2000; 2000US-0552317.  
 XX PR 09-JUL-2000; 2000US-0598042.  
 XX PR 19-JUL-2000; 2000US-0620312.  
 XX PR 03-AUG-2000; 2000US-0653450.  
 XX PR 14-SEP-2000; 2000US-0662191.  
 XX PR 19-OCT-2000; 2000US-0693036.  
 XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.  
 XX DR N-PSDB; AAI58944.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2933; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX SQ Sequence 258 AA;

Query Match 15.3%; Score 42; DB 22; Length 258;  
 Best Local Similarity 41.2%; Pred. No. 29;  
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 49 QGXRFXXXWKGVSVDV 65  
 : | | | | | | | | | |  
 Db 60 EGDEPITQWKGTVDQV 76

RESULT 15  
 AAM41573  
 ID AAM41573 standard; Protein: 279 AA.

XX AC AAM41573;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polypeptide SEQ ID NO 6504.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.  
 XX PR 25-APR-2000; 2000US-0552317.  
 XX PR 09-JUL-2000; 2000US-0598042.  
 XX PR 19-JUL-2000; 2000US-0620312.  
 XX PR 03-AUG-2000; 2000US-0653450.  
 XX PR 14-SEP-2000; 2000US-0662191.  
 XX PR 19-OCT-2000; 2000US-0693036.  
 XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI: 2001-442253/47.  
 XX DR N-PSDB; AAI60729.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -



DR WPI; 2001-514838/56.  
 XX N-PSDB; AAI84495.  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX PS  
 XX Claim 20; SEQ ID NO 18456; 1399pp + Sequence Listing: English.  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC - at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 106 AA;  
 SQ

Query Match 16.0%; Score 44; DB 22; Length 106;  
 Best Local Similarity 38.5%; Pred. No. 4.2;  
 Matches 10; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

QY 45 MXXQGXRXFXFXKXGVSVDVXGVFL 70  
 I I I I I I I I I I I I I I I I  
 Db 64 MLGVQGASRFGVGGVSWSIDGAVL 89

RESULT 10  
 ABB48287  
 ID ABB48287 standard; Protein: 434 AA.  
 XX  
 XX ABB48287;  
 XX  
 XX 05-FEB-2002 (first entry)  
 XX  
 XX Listeria monocytogenes protein #991.  
 DE  
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KW vitamin B12; bacterial infection; disease.  
 KW  
 XX Listeria monocytogenes.  
 OS  
 XX WO200177335-A2.  
 PN  
 XX 18-OCT-2001.  
 PD  
 XX  
 XX 11-APR-2001; 2001WO-FR01118.  
 PF  
 XX  
 XX 11-APR-2000; 2000FR-0004629.  
 PR  
 XX (INSP ) INST PASTEUR.  
 PA  
 XX Buchrieser C, Frangeul L, Couve E, Rusnok C, Fsihi H, Dehoux P;  
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Donann E, Hain T, Berche P, Charbit A, Durand L;  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 XX WPI; 2002-010914/01.  
 DR  
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT

PT related polypeptides -  
 XX  
 XX Claim 6; SEQ ID NO 992; 192pp; French.  
 XX  
 CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGB-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies. Identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 434 AA;  
 SQ

Query Match 15.6%; Score 43; DB 23; Length 434;  
 Best Local Similarity 57.1%; Pred. No. 36;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 62 VDSVXGVFLTQNXD 75  
 I I I I I I I I I I I I I I I I  
 Db 386 VDTGGWFLTQNYE 399

RESULT 11  
 AAU42256  
 ID AAU42256 standard; Protein: 436 AA.  
 XX  
 XX AAU42256;  
 XX  
 XX 27-FEB-2002 (first entry)  
 XX  
 XX Propionibacterium acnes immunogenic protein #3152.  
 DE  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 XX Propionibacterium acnes.  
 OS  
 XX WO200181581-A2.  
 PN  
 XX 01-NOV-2001.  
 PD  
 XX  
 XX 20-APR-2001; 2001WO-US12865.  
 PF  
 XX  
 XX 21-APR-2000; 2000US-199047P.  
 PR  
 XX 02-JUN-2000; 2000US-208841P.  
 PR  
 XX 07-JUL-2000; 2000US-216747P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 XX WPI; 2001-616774/71.  
 DR  
 XX N-PSDB; AAS59516.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX





CC region of DMT, a host cell comprising an exogenous polynucleotide  
CC encoding a DMT-like protein and a transgenic plant comprising a  
CC polynucleotide encoding a DMT-like protein. The expression cassette is  
CC useful for modulating transcription. The method comprises introducing  
CC the cassette into a host cell preferably Agrobacterium by sexual  
CC cross, and selecting a host cell with modulated transcription, where  
CC the protein is capable of exhibiting at least one of the following  
CC biological activities, which include enhanced expression of the  
CC protein in a plant results in a delay in flowering time, introduction of  
CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expression of the  
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.  
XX  
SQ Sequence 1729 AA;  
  
Query Match 59.6%; Score 164; DB 23; Length 1729;  
Best Local Similarity 43.0%; Pred. No. 1.3e-20;  
Matches 43; Conservative 0; Mismatches 47; Indels 10; Gaps 2;  
  
QY 1 KVVXDXXTXXWXXLXXXXXX-----XXERXXFXRXXXFXRMXXQGX 51  
DB 697 KVDIDETTRINLWLMGKGEKDEKDKKEKWEERVRFRGRADSFIRMHLVQGD 756  
  
QY 52 RXFXKMGVSDVSGVFLTON-XDXXSSXAXMXXAXFP 90  
DB 757 RRFSPKGSVDSVIGVFLTONVSDHLSAFAFMSLAARFP 796  
  
RESULT 4  
AAU72737  
ID AAU72737 standard; Protein; 1413 AA.  
XX  
AC AAU72737;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Arabidopsis DMT2 (1DMT2) protein.  
XX  
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX  
OS Arabidopsis Thaliana.  
XX  
PN WO200180626-A1.  
XX  
PD 01-NOV-2001.  
XX  
XX 23-APR-2001; 2001WO-US13059.  
XX  
XX 21-APR-2000; 2000US-0553690.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
PA Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
PI  
XX WPI; 2002-055307/07.  
DR N-PSDB; AAS96692.  
XX  
XX New polynucleotide that control plant development comprising a sequence  
XX having a specific homology to DEMETER domains A,B or C -  
XX  
XX Disclosure; Page 67; 109pp; English.  
XX  
XX The invention relates to an isolated polynucleotide sequence or their  
XX complement encoding a polypeptide having a sequence at least 40%  
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
XX or C or their combinations. Also included are an expression

CC cassette comprising the polynucleotide or comprising a heterologous  
CC polynucleotide under the control of a promoter at least 70% identical to  
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
CC region of DMT, a host cell comprising an exogenous polynucleotide  
CC encoding a DMT-like protein and a transgenic plant comprising a  
CC polynucleotide encoding a DMT-like protein. The expression cassette is  
CC useful for modulating transcription. The method comprises introducing  
CC the cassette into a host cell preferably Agrobacterium by sexual  
CC cross, and selecting a host cell with modulated transcription, where  
CC the protein is capable of exhibiting at least one of the following  
CC biological activities, which include enhanced expression of the  
CC protein in a plant results in a delay in flowering time, introduction of  
CC the protein into a cell results in modulation of methylation of  
CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expression of the  
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.  
XX  
SQ Sequence 1413 AA;  
  
Query Match 59.5%; Score 163.5; DB 23; Length 1413;  
Best Local Similarity 43.4%; Pred. No. 1.2e-20;  
Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;  
  
QY 1 KVVXDXXTXXWXXL-----XXXDXXXXXXXERXXFXRXMXXQGX 52  
DB 528 KVDLDETRVWKLLENINSEGVDSDEQAKWEEERNVFRGRADSFIRMHLVQGD 587  
  
QY 53 XFXKMGVSDVSGVFLTON-XDXXSSXAXMXXAXFP 90  
DB 588 RFPKGSVDSVGVFLTONVSDHLSAFAFMSLASQFP 626  
  
RESULT 5  
AAU72743  
ID AAU72743 standard; Protein; 1952 AA.  
XX  
AC AAU72743;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Rice DMT1 protein.  
XX  
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX  
OS Oryza sativa.  
XX  
PN WO200180626-A1.  
XX  
PD 01-NOV-2001.  
XX  
XX 23-APR-2001; 2001WO-US13059.  
XX  
XX 21-APR-2000; 2000US-0553690.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
PA Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
PI  
XX WPI; 2002-055307/07.  
DR N-PSDB; AAS96695.  
XX  
XX New polynucleotide that control plant development comprising a sequence  
XX having a specific homology to DEMETER domains A,B or C -  
XX  
XX Disclosure; Page 79-80; 109pp; English.  
XX  
XX The invention relates to an isolated polynucleotide sequence or their





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:08 ; Search time 13.9706 Seconds  
(without alignments)  
858.415 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KVXXDXXTXWXXLXXXXD.....TQNXDXSSXAMXAXXFP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166.5	60.5	1309	23 ABB92260	Herbicideally activ
2	166.5	60.5	1332	23 AAU72739	Arabidopsis DMT3 (
3	164	59.6	1729	23 AAU72736	Arabidopsis Demete
4	163.5	59.5	1413	23 AAU72737	Arabidopsis DMT2 (
5	162.5	59.1	1952	23 AAU72743	Rice DMT1 protein.
6	162	58.9	1114	23 AAU72741	Arabidopsis DMT4 (
7	48	17.5	235	22 AAU46627	Propionibacterium
8	46	16.7	882	22 AAB93159	Human protein sequ
9	44	16.0	106	22 AAO04564	Human polypeptide
10	43	15.6	434	23 ABB48287	Listeria monocytog

11	43	15.6	436	22 AAU42256	Propionibacterium
12	43	15.6	466	21 AAY95051	Candida albicans p
13	42	15.3	232	22 AAM39787	Human polypeptide
14	42	15.3	258	22 AAM39788	Human polypeptide
15	42	15.3	279	22 AAM41573	Human polypeptide
16	42	15.3	279	22 AAM41574	Human polypeptide
17	42	15.3	370	22 AAB93450	Human protein sequ
18	42	15.3	370	23 AAB47993	Pristanoyl-CoA oxi
19	42	15.3	973	19 AAU50778	Anino acid sequenc
20	41.5	15.1	351	22 AAB93450	Propionibacterium
21	41	14.9	331	22 ABG10353	Novel human diagno
22	41	14.9	453	22 ABG12472	Novel human diagno
23	41	14.9	973	23 ABB97191	Novel human protel
24	41	14.9	973	23 ABB05701	Human intracellula
25	40.5	14.7	84	20 AAY11870	Human 5' EST secre
26	40.5	14.7	310	23 ABB92345	Herbicideally activ
27	40	14.5	159	23 ABP38906	Staphylococcus epi
28	40	14.5	196	23 ABP30378	Streptococcus poly
29	40	14.5	200	23 ABP28344	Streptococcus poly
30	40	14.5	211	21 AAY51682	Murine clodin 7 pr
31	40	14.5	313	21 AAY90872	Human G protein-co
32	40	14.5	313	22 AAU10306	G-protein coupled
33	40	14.5	313	22 AAG72077	Human olfactory re
34	40	14.5	318	20 AAW90108	Human FLAME-2 prot
35	40	14.5	318	20 AAW90109	Mouse FLAME-2 prot
36	40	14.5	318	21 AAY51022	Human DEDD protein
37	40	14.5	318	21 AAY51023	Murine DEDD protei
38	40	14.5	318	22 AAB93016	Human protein sequ
39	40	14.5	318	22 AAB94040	Human protein sequ
40	40	14.5	1140	22 AAE09365	Human ATP-binding
41	40	14.5	1182	22 AAM93387	Human polypeptide,
42	40	14.5	1503	20 AAY43544	A human MPR-relate
43	40	14.5	1503	22 AAE09361	Human ATP-binding
44	40	14.5	1503	22 AAE09363	Human ATP-binding
45	40	14.5	1503	22 AAE09364	Human ATP-binding

ALIGNMENTS

RESULT 1	
ABB92260	
ID	ABB92260 standard; Protein; 1309 AA.
XX	
AC	ABB92260;
XX	
DT	31-MAY-2002 (first entry)
XX	
DE	Herbicideally active polypeptide SEQ ID NO 1471.
XX	
KW	Herbicideal; plant; agriculture; herbicide.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200210210-A2.
XX	
PD	07-FEB-2002.
XX	
PF	
XX	28-AUG-2001; 2001WO-EP09892.
XX	
PR	28-AUG-2001; 2001WO-EP09892.
XX	
PA	(FARB ) BAYER AG.
XX	
PI	Tietjen K, Weidler M;
XX	
DR	WPI; 2002-269010/31.
XX	
PT	Identifying plant target proteins for herbicideally active compounds,
PT	comprising aligning and comparing nucleic acid or amino acid sequences
PT	from plant with nucleic acid or amino acid sequences from non-plant
PT	organisms -
XX	

CITY: HILLSBOROUGH  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,236  
FILING DATE: March 11, 1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: EXEL98-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 343-4341  
TELEFAX: (650) 343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 634 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-041-236-2

Query Match 13.8%; Score 38; DB 4; Length 634;  
Best Local Similarity 58.3%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 50 GRRXFXWKGSV 61  
| | | | |  
Db 19 GPRIFAVWKGHV 30

## RESULT 15

US-09-240-410-2  
Sequence 2, Application US/09240410  
Patent No. 6197544  
GENERAL INFORMATION:  
APPLICANT: MICHALOVICH, DAVID  
APPLICANT: HAYES, PHILIP DAVID  
TITLE OF INVENTION: NOVEL COMPOUNDS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ratner & Prestia  
STREET: P.O. Box 980  
CITY: Valley Forge  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/240,410  
FILING DATE: 27-JAN-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: UK APPLICATION NO. TO BE ASSIGNED  
FILING DATE: 20-JAN-1999  
APPLICATION NUMBER: EP APPLICATION NO. 98300694.1  
FILING DATE: 30-JAN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Prestia, Paul F  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GP-30039  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700

TELEFAX: 610-407-0700  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 666 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-240-410-2

Query Match 13.8%; Score 38; DB 4; Length 666;  
Best Local Similarity 58.3%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 50 GRRXFXWKGSV 61  
| | | | |  
Db 51 GPRIFAVWKGHV 62

Search completed: November 8, 2002, 19:29:14  
Job time : 7.14706 secs

;; GENERAL INFORMATION:  
;; APPLICANT: Garner, Ian  
;; APPLICANT: Dailymple, Michael A  
;; APPLICANT: Prunkard, Donna E  
;; APPLICANT: Foster, Donald C  
;; TITLE OF INVENTION: Production of Fibrinogen in Transgenic  
;; TITLE OF INVENTION: Animals  
;; NUMBER OF SEQUENCES: 27  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: ZymoGenetics, Inc.  
;; STREET: 4225 Roosevelt Way, N.E.  
;; CITY: Seattle  
;; STATE: WA  
;; COUNTRY: USA  
;; ZIP: 98105  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/206.176  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Parker, Gary E  
;; REGISTRATION NUMBER: 31-648  
;; REFERENCE/DOCKET NUMBER: 93-15  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 206-547-8080 ext 322  
;; TELEFAX: 206-548-2329  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 491 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-206-176-4  
  
Query Match 14.2%; Score 39; DB 1; Length 491;  
Best Local Similarity 47.1%; Pred. No. 78;  
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
  
Qy 57 WKGSVDSVXGVFLTON 73  
||| | : | | | |  
Db 347 WKGDVKVKAHYGGFTVQN 363  
  
RESULT 12  
US-09-620-412C-317  
; Sequence 317, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620.412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 317  
; LENGTH: 646  
; TYPE: PRT  
; ORGANISM: Chlamydia trachomatis  
US-09-620-412C-317  
  
Query Match 14.2%; Score 39; DB 4; Length 646;  
Best Local Similarity 27.0%; Pred. No. 1.1e+02;  
Matches 10; Conservative 4; Mismatches 23; Indels 0; Gaps 0;  
  
Qy 37 RXXAFXBMXXQGXRFXXWKGSVDSVXGVFLTON 73  
| | | | | : | : | : | : |

Db 326 RTLAFINNMGGCGAISADTQISITDVKILFENN 362  
  
RESULT 13  
US-08-602-359A-33  
; Sequence 33, Application US/08602359A  
; Patent No. 5942430  
; GENERAL INFORMATION:  
; APPLICANT: ROBERTSON, Daniel E.  
; APPLICANT: MURPHY, Dennis  
; APPLICANT: REID, John  
; APPLICANT: MAFFIA, Anthony  
; APPLICANT: LINK, Steven  
; APPLICANT: SWANSON, Ronald V.  
; APPLICANT: WARREN, Patrick V.  
; APPLICANT: KOSMOTKA, Anna  
; TITLE OF INVENTION: ESTERASES  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & RICHARDSON P.C.  
; STREET: 4225 EXECUTIVE SQUARE, STE 1400  
; CITY: LA JOLLA  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/602.359A  
; FILING DATE: February 16, 1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: HAILE, LISA A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/010001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-678-5070  
; TELEFAX: 619-678-5099  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 184 AMINO ACIDS  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-602-359A-33  
  
Query Match 13.8%; Score 38; DB 2; Length 184;  
Best Local Similarity 45.8%; Pred. No. 37;  
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 2;  
  
Qy 50 GXRFXFXW---KGSVDSVXGVFL 70  
| | | | | : | : | : | : |  
Db 72 GLRNLWSVRVKGSLIISM-GVFL 94  
  
RESULT 14  
US-09-041-236-2  
; Sequence 2, Application US/09041236  
; Patent No. 6225285  
; GENERAL INFORMATION:  
; APPLICANT: Luo, Yuling  
; APPLICANT: Xiomel, Xu  
; TITLE OF INVENTION: Semaphorin K1  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 75 DENISE DRIVE

```
RESULT 8
US-09-276-993-4
; Sequence 4, Application US/09276993
; Patent No. 6207801
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Enad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6207801Iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-4
;
; Query Match 14.5%; Score 40; DB 4; Length 318;
; Best Local Similarity 33.3%; Pred. No. 31;
; Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;
;
Qy 57 WK----GSVDSVXGVFTQN 73
I: ||::|||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 9
US-09-276-993-6
; Sequence 6, Application US/09276993
; Patent No. 6207801
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Enad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6207801Iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-6
;
; Query Match 14.5%; Score 40; DB 4; Length 318;
; Best Local Similarity 33.3%; Pred. No. 31;
; Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;
;
Qy 57 WK----GSVDSVXGVFTQN 73
I: ||::|||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 9
US-09-276-993-6
; Sequence 6, Application US/09276993
; Patent No. 6207801
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Enad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6207801Iris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-6
;
; Query Match 14.2%; Score 39; DB 4; Length 398;
; Best Local Similarity 36.8%; Pred. No. 61;
; Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
;
Qy 52 RFXFXWKSGSVDSVXGVFL 70
I: |::|:|:|
Db 22 RIWKTWRNSIKDNVCGEIL 40

RESULT 11
US-08-206-176-4
; Sequence 4, Application US/08206176
; Patent No. 5639940
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,993
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/859,167
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-276-993-6
;
; Query Match 14.5%; Score 40; DB 4; Length 318;
; Best Local Similarity 33.3%; Pred. No. 31;
; Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;
;
Qy 57 WK----GSVDSVXGVFTQN 73
I: ||::|||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 10
US-09-213-053-3
; Sequence 3, Application US/09213053
; Patent No. 6159477
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe
; APPLICANT: BAUDU, Philippe
; TITLE OF INVENTION: CANINE HERPESVIRUS BASED RECOMBINANT LIVE VACCINE, IN
; TITLE OF INVENTION: PARTICULAR AGAINST CANINE DISTEMPER, RABIES OR THE
; TITLE OF INVENTION: PARAINFLUENZA 2 VIRUS
; FILE REFERENCE: 454313-2200
; CURRENT APPLICATION NUMBER: US/09/213,053
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 9608242
; EARLIER FILING DATE: 1996-06-27
; EARLIER APPLICATION NUMBER: PCT/FR97/01115
; EARLIER FILING DATE: 1997-06-23
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Canine herpesvirus
; US-09-213-053-3
;
; Query Match 14.2%; Score 39; DB 4; Length 398;
; Best Local Similarity 36.8%; Pred. No. 61;
; Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
;
Qy 52 RFXFXWKSGSVDSVXGVFL 70
I: |::|:|:|
Db 22 RIWKTWRNSIKDNVCGEIL 40

RESULT 11
US-08-206-176-4
; Sequence 4, Application US/08206176
; Patent No. 5639940
```

STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: WINDOWS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/859,167  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 318 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-859-167-6

Query Match 14.5%; Score 40; DB 3; Length 318;  
Best Local Similarity 33.3%; Pred. No. 31;  
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVDSVXGVFLTN 73  
I: ||:||||: |||:|  
Db 260 WRDYINGSLLLEALKGVFTDS 280

RESULT 6  
US-09-109-273-4  
Sequence 4, Application US/09109273  
Patent No. 6063760  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
APPLICANT: Fernandez-Alnemri, Teresa  
TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF  
TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS  
TITLE OF INVENTION: OF MAKING THE SAME  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6063760ris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: WINDOWS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/109,273  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/859,167  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 318 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-109-273-4

Query Match 14.5%; Score 40; DB 3; Length 318;  
Best Local Similarity 33.3%; Pred. No. 31;  
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVDSVXGVFLTN 73  
I: ||:||||: |||:|  
Db 260 WRDYINGSLLLEALKGVFTDS 280

RESULT 7  
US-09-109-273-6  
Sequence 6, Application US/09109273  
Patent No. 6063760  
GENERAL INFORMATION:  
APPLICANT: Alnemri, Emad S.  
APPLICANT: Fernandez-Alnemri, Teresa  
TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF  
TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS  
TITLE OF INVENTION: OF MAKING THE SAME  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6063760ris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: WINDOWS  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/109,273  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/859,167  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: TJU-  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 318 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-109-273-6

Query Match 14.5%; Score 40; DB 3; Length 318;  
Best Local Similarity 33.3%; Pred. No. 31;  
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK----GSVDSVXGVFLTN 73  
I: ||:||||: |||:|  
Db 260 WRDYINGSLLLEALKGVFTDS 280

```

US-09-517-639-4
; Sequence 4, Application US/09517639
; Patent No. 6414120
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo M.
; TITLE OF INVENTION: Mammalian Deep Orange Proteins
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,639
; FILING DATE:
; CLASSIFICATION:
; * PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,452
; FILING DATE: 31-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1240.004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-517-639-4

```

```

RESULT 3
US-09-134-001C-3751
; Sequence 3751, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYNN Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3751
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3751

```

```

Query Match          14.5%; Score 40; DB 4; Length 159;
Best Local Similarity 24.3%; Pred. No. 14;
Matches 9; Conservative 5; Mismatches 23; Indels 0; Gaps

QY 54 FXXWKGSVDSVGVFLTQXNDXXSSXAXMXXXXP 90
      | || | : | : | : | : | : | : |
Db 92 FATSKGKDSMPGMHITLNFEEVKGKTTVTSTSTP 128

RESULT 4
US-08-859-167-4
; Sequence 4, Application US/08859167
; Patent No. 6037461
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS OF
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6037461ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/859,167
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Defuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 318 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-859-167-4

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 5.14706 Seconds  
(without alignments)  
514.481 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KVVXDXXTXXXXXXLXXXXX.....TQNXDXSSXAMXXXXXFP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pap.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	15.3	973	3	US-08-904-452-4
2	42	15.3	973	4	US-09-517-639-4
3	40	14.5	159	4	US-09-134-001C-3751
4	40	14.5	318	3	US-08-859-167-4
5	40	14.5	318	3	US-08-859-167-6
6	40	14.5	318	3	US-09-109-273-4
7	40	14.5	318	3	US-09-109-273-6
8	40	14.5	318	4	US-09-276-993-4
9	40	14.5	318	4	US-09-276-993-6
10	39	14.2	398	4	US-09-213-053-3
11	39	14.2	491	1	US-08-206-176-4
12	39	14.2	646	4	US-09-620-412C-317
13	38	13.8	184	2	US-08-602-359A-33
14	38	13.8	634	4	US-09-041-236-2
15	38	13.8	666	4	US-09-240-410-2
16	37.5	13.6	1198	4	US-09-199-637A-405
17	37	13.5	138	2	US-09-057-762-13
18	37	13.5	138	3	US-08-326-119A-13
19	37	13.5	271	3	US-08-433-263B-2
20	37	13.5	301	4	US-09-134-001C-4909
21	37	13.5	1127	4	US-08-937-195-3
22	37	13.5	1127	4	US-08-915-152-3
23	37	13.5	15281	5	PCT-US96-07627-3
24	37	13.5	15281	2	US-08-471-119A-2
25	36.5	13.3	175	4	US-08-858-207A-363
26	36.5	13.3	245	2	US-08-715-204-3
27	36.5	13.3	245	3	US-09-162-597-3

28 36 13.1 1213 2 US-08-937-102-2 Sequence 2, Appli  
29 36 13.1 1443 2 US-08-670-707A-39 Sequence 39, Appli  
30 36 13.1 1443 4 US-09-037-601-39 Sequence 39, Appli  
31 36 13.1 1443 4 US-09-315-179-39 Sequence 39, Appli  
32 36 13.1 1467 4 US-09-523-656-38 Sequence 38, Appli  
33 36 13.1 2115 4 US-09-324-867-5 Sequence 5, Appli  
34 36 13.1 2133 2 US-08-670-707A-37 Sequence 37, Appli  
35 36 13.1 2133 4 US-09-037-601-37 Sequence 37, Appli  
36 36 13.1 2133 4 US-09-315-179-37 Sequence 30, Appli  
37 36 13.1 2133 4 US-09-523-656-30 Sequence 37, Appli  
38 36 13.1 3898 2 US-08-876-991-2 Sequence 2, Appli  
39 36 13.1 3898 2 US-09-059-853-2 Sequence 2, Appli  
40 35.5 12.9 365 4 US-09-151-592-2 Sequence 2, Appli  
41 35.5 12.9 365 4 US-09-254-077A-10 Sequence 2, Appli  
42 35 12.7 90 4 US-08-936-165A-270 Sequence 10, Appli  
43 35 12.7 108 4 US-09-134-001C-3747 Sequence 270, App  
44 35 12.7 161 4 US-09-413-814-15 Sequence 3747, Ap  
45 35 12.7 184 1 US-08-075-193-6 Sequence 15, Appli  
Sequence 6, Appli

## ALIGNMENTS

RESULT 1  
US-08-904-452-4  
; Sequence 4, Application US/08904452  
; Patent No. 6083742  
; GENERAL INFORMATION:  
; APPLICANT: Randazzo, Filippo M.  
; TITLE OF INVENTION: Mammalian Deep Orange Proteins  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/904.452  
; FILING DATE: 31-JUL-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Potter, Jane E.R.  
; REGISTRATION NUMBER: 33,332  
; REFERENCE/DOCKET NUMBER: 1240.004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-2718  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 973 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-904-452-4

Query Match 15.3%; Score 42; DB 3; Length 973;  
Best Local Similarity 47.1%; Pred. No. 51;  
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 49 QGXRFXWKGVSVDV 65

Db 129 OKARPLARKGOLVESV 145

RESULT 2



**THIS PAGE BLANK (USPTO)**

```
QY 67 GVFLTQNDXXSSXAMXXA 86
      | : | | | | | | :
Db 86 GIFTTPEDETSSATTISTS 105

RESULT 13
US-09-836-077-3
; Sequence 3, Application US/09836077
; Patent No. US20020037851A1
; GENERAL INFORMATION:
; APPLICANT: FLECKENSTEIN, Bernhard
; ENSER, Armin
; TITLE OF INVENTION: HUMAN SEMAPHORIN L (H-SEMAL) AND
; CORRESPONDING SEMAPHORINS IN OTHER SPECIES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Frommer Lawrence & Haug LLP
; STREET: 745 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/836,077
; FILING DATE: 16-Apr-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lawrence, William F.
; REGISTRATION NUMBER: 28,029
; REFERENCE/DOCKET NUMBER: 514429-3647
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 amino acids
; TYPE: amino acid
; STRANDEDNESS: n/a
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-836-077-3
Query Match 13.8%; Score 38; DB 10; Length 666;
Best Local Similarity 58.3%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 50 GXRFXFXWKGSV 61
      | : | | | | | | :
Db 51 GPRIFAVWKGHV 62

RESULT 14
US-09-764-587A-2
; Sequence 2, Application US/09764587A
; Patent No. US20020106722A1
; GENERAL INFORMATION:
; APPLICANT: David Michalovich
; APPLICANT: Philip David Hayes
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30039-D1
; CURRENT APPLICATION NUMBER: US/09/764,587A
; CURRENT FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 09/240,410
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: EP 98300694.1
; PRIOR FILING DATE: 1999-01-30
```

```
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 666
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-764-587A-2
Query Match 13.8%; Score 38; DB 10; Length 666;
Best Local Similarity 58.3%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 50 GXRFXFXWKGSV 61
      | : | | | | | | :
Db 51 GPRIFAVWKGHV 62

RESULT 15
US-09-799-875-2
; Sequence 2, Application US/09799875
; Patent No. US20020034780A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: No. US20020034780A1 Human Protein Kinases and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: 35800/209996
; CURRENT APPLICATION NUMBER: US/09/799,875
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/182,059
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/659,287
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-799-875-2
Query Match 13.8%; Score 38; DB 10; Length 879;
Best Local Similarity 30.3%; Pred. No. 72;
Matches 10; Conservative 2; Mismatches 21; Indels 0; Gaps 0;

QY 58 KGSVVDVXGVFLTQNDXXSSXAMXXAXFP 90
      | : | | | | | | :
Db 542 KNSKRDYLSSTFLCSDDDDRASKNISMNSDSSFP 574

Search completed: November 8, 2002, 19:29:40
Job time : 4.79412 secs
```

```
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-317

Query Match 14.2%; Score 39; DB 10; Length 646;
Best Local Similarity 27.0%; Pred. No. 34;
Matches 10; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

Qy 37 RXXFXRMXXQGXRFXXKXGVSVDVXGVFELTON 73
| | | | | | | | | | | | | | | | | | | |
Db 326 RFLAFINMSGDCGGAISADTQISTITVKGILFENN 362

RESULT 10
US-09-903-410-33
; Sequence 33, Application US/09903410
; Patent No. US20020146799A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: ROBERTSON, Dan
; APPLICANT: MURPHY, Dennis
; APPLICANT: REID, John
; APPLICANT: MAFFIA, Anthony
; APPLICANT: LINK, Steven
; APPLICANT: SWANSON, Ronald
; APPLICANT: WARREN, Patrick
; APPLICANT: KOSMOTKA, Anna
; TITLE OF INVENTION: ENZYMES HAVING ESTERASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: DIVER1180-2
; CURRENT APPLICATION NUMBER: US/09/903,410
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 09/382,242
; PRIOR FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/602,359
; PRIOR FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Staphylothermus Marinus
US-09-903-410-33

Query Match 13.8%; Score 38; DB 10; Length 184;
Best Local Similarity 45.8%; Pred. No. 12;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 2;

Qy 50 GXRFXFXW---KGSVDSVXGVFL 70
| | | | | | | | | | | | | | | | | | | |
Db 72 GLRNLWSVRKGLISM-GVFL 94

RESULT 11
US-09-935-390A-36
; Sequence 36, Application US/09935390A
; Patent No. US200200761A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; Quianjin, Hu
; Garcia, Pablo
; Williams, Lewis T.
; Kothakota, Srinivas
; TITLE OF INVENTION: Secreted Human Proteins
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; City: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/935,390A
; FILING DATE: 22-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/988,671
; FILING DATE: 1997-12-11
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. R. Potter
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1369.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2718
; TELEFAX: (510) 655-3542
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. US200200761A1e
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-935-390A-36

Query Match 13.8%; Score 38; DB 10; Length 210;
Best Local Similarity 24.4%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 26; Indels 0; Gaps 0;

Qy 50 GXRFXFXKXGVSVDVXGVFELTONXXSSXAXMXAXXFP 90
| | | | | | | | | | | | | | | | | | | |
Db 162 GPAFIGWAGSALVILGALLSCSCPGNESKAGYAPRSYP 202

RESULT 12
US-09-764-864-1235
; Sequence 1235, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1235
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (14)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (488)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (493)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (494)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1235

Query Match 13.8%; Score 38; DB 10; Length 494;
Best Local Similarity 35.0%; Pred. No. 37;
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
```

```

; PRIOR APPLICATION NUMBER: 60/178,227
; PRIOR FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 60/220,590
; PRIOR FILING DATE: 2000-07-25
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 86
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-761-288-86

Query Match          14.2%; Score 39; DB 10; Length 305;
Best Local Similarity 38.9%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 54 FXXWKGVSVDVXGVFLT 71
      | | | : | : |||
DB 146 FALWLGGTTHSLVQTFLT 163

RESULT 8
US-09-825-414-56
; Sequence 56, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 56
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. delphinii
US-09-825-414-56

Query Match          14.2%; Score 39; DB 10; Length 316;
Best Local Similarity 53.8%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 58 KGSVDSVXGVFL 70
      || : || | |||
DB 101 KGQIDLVSGAFL 113

RESULT 9
US-09-841-132-317
; Sequence 317, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 317

```

```
; TITLE OF INVENTION: Protein for Regulation of Apoptosis
; FILE REFERENCE: 4121-120
; CURRENT APPLICATION NUMBER: US/09/733.167
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: PCT/DE99/01712
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: German Patent Application No. US20020099009A1 198 25 621.3
; PRIOR FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-733-167-1

Query Match          14.5%; Score 40; DB 10; Length 318;
Best Local Similarity 33.3%; Pred. No. 9.9;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK---GSVVDVSVXGVFLTON 73
      |||::: |||:|
Db - 260 WRDYINGSLLEALKGVFTDS 280

RESULT 3
US-09-733-167-3
; Sequence 3, Application US/09733167
; Patent No. US20020099009A1
; GENERAL INFORMATION:
; APPLICANT: Peter, Marcus
; APPLICANT: Krammer, Peter
; TITLE OF INVENTION: Protein for Regulation of Apoptosis
; FILE REFERENCE: 4121-120
; CURRENT APPLICATION NUMBER: US/09/733.167
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: PCT/DE99/01712
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: German Patent Application No. US20020099009A1 198 25 621.3
; PRIOR FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-733-167-3

Query Match          14.5%; Score 40; DB 10; Length 318;
Best Local Similarity 33.3%; Pred. No. 9.9;
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

Qy 57 WK---GSVVDVSVXGVFLTON 73
      |||::: |||:|
Db 260 WRDYINGSLLEALKGVFTDS 280

RESULT 4
US-09-815-242-13509
; Sequence 13509, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
```

```
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13509
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13509

Query Match          14.4%; Score 39.5; DB 10; Length 164;
Best Local Similarity 37.0%; Pred. No. 5.7;
Matches 10; Conservative 4; Mismatches 10; Indels 3; Gaps 1;

Qy 58 KGSVVD---SVXGVFLTONDXDXSSXA 81
      ||| | | | | | | | | | | | | | |
Db 126 KGEVADLTISLAGKIISQNLDSHAHKA 152

RESULT 5
US-09-815-242-13630
; Sequence 13630, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13630
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13630
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:26:04 ; Search time 2.79412 Seconds  
(without alignments)  
464.428 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KXXXXXXXXXXX.....TQNXSSXAMXXXXXP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
- score greater than or equal to the score of the result being printed,  
- and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	14.5	210	US-09-733-167-6	Sequence 6, Appli
2	40	14.5	318	US-09-733-167-1	Sequence 1, Appli
3	40	14.5	318	US-09-733-167-3	Sequence 3, Appli
4	39.5	14.4	164	US-09-815-242-13509	Sequence 13509, A
5	39.5	14.4	164	US-09-815-242-13630	Sequence 13630, A
6	39	14.2	241	US-09-825-414-22	Sequence 22, Appli
7	39	14.2	305	US-09-761-288-86	Sequence 86, Appli
8	39	14.2	316	US-09-825-414-56	Sequence 56, Appli
9	39	14.2	646	US-09-841-132-317	Sequence 317, App
10	38	13.8	184	US-09-903-410-33	Sequence 33, Appli
11	38	13.8	210	US-09-935-390A-36	Sequence 36, Appli
12	38	13.8	494	US-09-764-864-1235	Sequence 1235, Ap
13	38	13.8	666	US-09-836-077-3	Sequence 3, Appli
14	38	13.8	666	US-09-764-587A-2	Sequence 2, Appli
15	38	13.8	879	US-09-799-875-2	Sequence 2, Appli
16	37.5	13.6	167	US-09-815-242-13543	Sequence 13543, A
17	37.5	13.6	792	US-09-995-587A-11	Sequence 11, Appli
18	37	13.5	217	US-09-799-777-26	Sequence 26, Appli
19	37	13.5	276	US-09-815-242-11729	Sequence 11729, A

20	37	13.5	453	10	US-09-864-761-37331	Sequence 37331, A
21	36.5	13.3	158	12	US-10-052-586-6	Sequence 6, Appli
22	36	13.1	36	10	US-09-864-761-43486	Sequence 43486, A
23	36	13.1	53	10	US-09-864-761-42021	Sequence 42021, A
24	36	13.1	272	10	US-09-815-242-11074	Sequence 11074, A
25	36	13.1	310	10	US-09-886-055-153	Sequence 153, App
26	36	13.1	479	9	US-09-992-598-216	Sequence 216, App
27	36	13.1	479	10	US-09-989-722-216	Sequence 216, App
28	36	13.1	479	10	US-09-989-723-216	Sequence 216, App
29	36	13.1	479	10	US-09-989-279-216	Sequence 216, App
30	36	13.1	479	10	US-09-989-727-216	Sequence 216, App
31	36	13.1	479	10	US-09-989-731-216	Sequence 216, App
32	36	13.1	479	10	US-09-989-732-216	Sequence 216, App
33	36	13.1	479	10	US-09-991-073-216	Sequence 216, App
34	36	13.1	479	10	US-09-990-442-216	Sequence 216, App
35	36	13.1	479	10	US-09-991-163-216	Sequence 216, App
36	36	13.1	479	10	US-09-993-604-216	Sequence 216, App
37	36	13.1	479	10	US-09-990-456-216	Sequence 216, App
38	36	13.1	479	10	US-09-989-721-216	Sequence 216, App
39	35.5	12.9	365	10	US-09-051-034A-4	Sequence 4, Appli
40	35.5	12.9	522	10	US-09-815-242-11002	Sequence 11002, A
41	35	12.7	59	10	US-09-864-761-41743	Sequence 41743, A
42	35	12.7	90	10	US-09-939-980-270	Sequence 270, App
43	35	12.7	99	10	US-09-911-826A-22	Sequence 22, Appli
44	35	12.7	118	10	US-09-816-248-1	Sequence 1, Appli
45	35	12.7	307	10	US-09-815-242-4958	Sequence 4958, Ap

#### ALIGNMENTS

RESULT 1  
US-09-733-167-6  
; Sequence 6, Application US/09733167  
; Patent No. US20020099009A1  
; GENERAL INFORMATION:  
; APPLICANT: Peter, Marcus  
; APPLICANT: Krammer, Peter  
; TITLE OF INVENTION: Protein for Regulation of Apoptosis  
; FILE REFERENCE: 4121-120  
; CURRENT APPLICATION NUMBER: US/09/733,167  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: PCT/DE99/01712  
; PRIOR FILING DATE: 1999-06-08  
; PRIOR APPLICATION NUMBER: German Patent Application No. US20020099009A1 198 25 621  
; PRIOR FILING DATE: 1998-06-08  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 210  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: C-DEDD, which is a deletion mutant of human DEDD comprising am  
US-09-733-167-6  
Query Match 14.5%; Score 40; DB 10; Length 210;  
Best Local Similarity 33.3%; Pred. No. 6.2;  
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;

QY 57 WK-----GSVDSVXGVFLTQN 73  
| : ||::: ||| :  
DB 152 WRDYINGSLLEALKGVFTDS 172

RESULT 2  
US-09-733-167-1  
; Sequence 1, Application US/09733167  
; Patent No. US20020099009A1  
; GENERAL INFORMATION:  
; APPLICANT: Peter, Marcus  
; APPLICANT: Krammer, Peter

RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=98447599; PubMed=9774341;  
RA Stegh A.H., Schickling O., Ehret A., Scaffidi C., Peterhaensel C.,  
RA Hofmann T.G., Grummt I., Krammer P.H., Peter M.E.;  
RT "DEDD, a novel death effector domain-containing protein, targeted to  
RT the nucleolus";  
RL EMBO J. 17:5974-5986(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
RC TISSUE=Testis;  
RX MEDLINE=99049260; PubMed=9832420;  
RA Leo C.P., Hsu S.Y., McGee E.A., Salanova M., Hsueh A.J.W.;  
RT "DEFT, a novel death effector domain-containing molecule predominantly  
RT expressed in testicular germ cells";  
RL Endocrinology 139:4839-4848(1998).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Pan G.;  
RT "FUED-1, a novel molecule with a DED-like domain";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Thome M., Tschopp J.;  
RT "DEDPOL, a novel DED-containing protein";  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RA TISSUE=Blood;  
RC Zhao Z., Huang X., Li N., Zhu X., Cao X.;  
RT "A novel gene from human dendritic cell";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA TISSUE=Bone marrow, and Placenta;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Binds strongly and non-specifically to DNA. Induces  
CC apoptosis. Inhibits DNA transcription in vitro (By  
CC similarity).  
CC -!- SUBUNIT: Interacts with caspase-8 and FADD (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Translocated to the nucleus  
CC during CD95-mediated apoptosis where it is localized in the  
CC nucleoli (By similarity).  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
CC produced by alternative splicing.  
CC -!- TISSUE SPECIFICITY: Widely expressed with highest levels in  
CC testis.  
CC -!- SIMILARITY: CONTAINS 1 DEATH EFFECTOR DOMAIN (DED).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF083236; AAC33105.1; -  
DR EMBL; AF100341; AAD16414.1; -  
DR EMBL; AF043733; AAC80280.1; -  
DR EMBL; AJ010973; CAQ09445.1; -  
DR EMBL; AF064605; AAC17110.3; -  
DR EMBL; BC016724; AAH16724.1; -  
DR EMBL; BC013910; AAH13910.1; -  
DR Genew; HGNC:2755; DEDD.  
DR MIM; 606841; -  
DR InterPro; IPR001875; DED.  
DR Pfam; PF01335; DED; 1.  
DR SMART; SM00031; DED; 1.  
DR PROSITE; PS0168; DED; 1.  
KW Apoptosis; Transcription regulation; Repressor; DNA-binding;  
KW Nuclear protein; Alternative splicing.  
FF . DOMAIN 25 103 DED.

FT VARSPLIC 194 194 D-> GEEIQGFQWRSLEGEYKELLGHWAIVAIOY (IN  
FT ISOFORM 2).  
FT CONFLICT 13 13 P-> L (IN REF. 5).  
SQ SEQUENCE 318 AA; 36794 MW; FF9D5FF9B61F6BB6 CRC64;  
Query Match 14.5%; Score 40; DB 1; Length 318;  
Best Local Similarity 33.3%; Pred. No. 24;  
Matches 7; Conservative 8; Mismatches 2; Indels 4; Gaps 1;  
QY 57 WK---GSVVDVSVXGVFLTN 73  
I: ||::: |||:|  
Db 260 WRDYINGSLLLEALKGVFITDS 280  
Search completed: November 8, 2002, 19:27:09  
Job time : 5.23529 secs

```
CC or send an email to license@isb-sib.ch).
DR -----
DR EMBL; AE000030; AAB95974.1; -.
DR HSSP; Q9KWU7; lHQM.
DR InterPro; IPR001572; RNA_pol_B.
DR Pfam; PF00562; RNA_pol_B; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
DR Transferrase; Transcription; DNA-directed RNA polymerase;
KW Complete proteome.
SQ SEQUENCE 1391 AA; 155621 MW; B2F345AB24F18EAD CRC64;

Query Match 14.9%; Score 41; DB 1; Length 1391;
Best Local Similarity 64.3%; Pred. No. 78;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 58 KGSVVDVSVXGVFLT 71
| | | | | | | |
Db 118 KNSATSSVDGVFLT 131

RESULT 13
CLD7_RAT
ID CLD7_RAT STANDARD; PRT; 191 AA.
AC Q921L1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Claudin-7 (Fragment).
GN CLDN7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN SEQUENCE FROM N.A.
RA Keen T.J., Inglehearn C.F.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ011811; CAA09790.1; -.
DR InterPro; IPR001832; Claudin.
DR Pfam; PF00822; PMP22_Claudin.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 1 1
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
SQ SEQUENCE 191 AA; 20366 MW; 83B445908DFF41A CRC64;

Query Match 14.5%; Score 40; DB 1; Length 191;
Best Local Similarity 24.4%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 26; Indels 0; Gaps 0;

Qy 50 GXRFXFXWKSVDVSVXGVFLTQNXDXSSXAXMXXAXXP 90
| | | | | : | : | : |
Db 142 GPAIFGWAGSALVLLGGALLSCSPGSKAAYRAPRSYP 182

RESULT 14
CLD7_MOUSE
ID DEDD_HUMAN STANDARD; PRT; 318 AA.
AC 075618; O60737;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Death effector domain-containing protein (death effector domain-
DE containing testicular molecule) (DEDpro1) (FLDED-1) (K505).
GN DEDD OR DEFT OR DEDPRO1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```



RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RL *subtilis*.";  
 CC Nature 390:249-256(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE GPF0053 FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y14082; CAA74500.1; -;  
 DR EMBL: Z99109; CAB12794.1; -;  
 DR Subtilisin; BG13022; yndP.  
 DR InterPro: IPR000644; CBS\_Domain.  
 DR InterPro: IPR005170; CorC\_HlyC.  
 DR - InterPro: IPR002550; DUF21.  
 DR Pfam: PF00571; CBS; 2.  
 DR Pfam: PF01595; DUF21; 1.  
 DR Pfam: PF03471; CorC\_HlyC; 1.  
 DR SMART; SM00116; CBS; 2.  
 DR SMART; SM00326; SH3; 1.  
 KW Hypothetical protein; CBS domain; Repeat; Transmembrane;  
 KW Complete proteome.  
 FT TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 61 81 POTENTIAL.  
 FT TRANSMEM 101 121 POTENTIAL.  
 FT TRANSMEM 101 121 POTENTIAL.  
 SQ SEQUENCE 444 AA; 49855 MW; E3A0F57ABE8FB323 CRC64;  
 -----  
 Query Match 14.9%; Score 41; DB 1; Length 444;  
 Best Local Similarity 57.1%; Pred. No. 24;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 OY 62 VDSVXGVFLTNQXD 75  
 DB 385 VDTGGVFLTNQYD 398  
 -----  
 RESULT 11  
 MYSC DICDI  
 ID MYSC DICDI STANDARD; PRT; 1181 AA.  
 AC P42522;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin IC heavy chain.  
 GN MYOC OR DMIC.  
 OS Dictyostelium discoideum (Slime mold).  
 CC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 CC NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RX MEDLINE=95348228; PubMed=7622596;  
 RA Peterson M.D., Novak K.D., Ready M.C., Ruman J.I., Titus M.A.;  
 RT "Molecular genetic analysis of myoc, a Dictyostelium myosin I.";  
 RL J. Cell Sci. 108:1093-1103(1995).  
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE  
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.  
 CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED, DIMER OF A HEAVY  
 CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L35323; AAC37427.1; -;  
 DR HSPSP; P08799; IMND.  
 DR DictyDb; DD01090; myoc.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00018; SH3; 1.  
 DR Pfam: PF00063; myosin\_head; 2.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR ProDom; PD000066; SH3; 1.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00242; MYSc; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Myosin; Actin-binding; ATP-binding; Chemotaxis; SH3 domain;  
 KW Multigene family.  
 FT DOMAIN 1 ? MYOSIN HEAD-LIKE.  
 FT DOMAIN ? 1181 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.  
 FT NP\_BIND 109 116 ATP (POTENTIAL).  
 FT DOMAIN 1122 1181 SH3  
 SQ SEQUENCE 1181 AA; 132915 MW; 5EB1EE47FOCA8803 CRC64;  
 -----  
 Query Match 14.9%; Score 41; DB 1; Length 1181;  
 Best Local Similarity 31.6%; Pred. No. 66;  
 Matches 6; Conservative 4; Mismatches 9; Indels 0; Gaps 0;  
 OY 57 WGSVDSVXGVFLTNQXD 75  
 DB 653 WKGDAISACRAILASQNVND 671  
 -----  
 RESULT 12  
 RPOB\_MYCPN  
 ID RPOB\_MYCPN STANDARD; PRT; 1391 AA.  
 AC P78013;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase  
 DE beta chain) (RNA polymerase beta subunit).  
 GN RPOB OR MPN516 OR MP326.  
 OS Mycoplasma pneumoniae.  
 CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 CC NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
 RA Herrmann R.;  
 RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma*  
 RT *pneumoniae*.";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC {RNA}(N).  
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

```

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z98682; CAB11357.1; -;
CC EMBL: Z99111; CAB13377.1; -;
CC Subtilist; BG13363; ylbk.
CC InterPro; IPR002641; Patatin.
CC Pfam; PF01734; Patatin; 1
CC PROSITE; PS01237; UPF0028; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 260 AA; 28258 MW; 363B6DDF017BE77E CRC64;

Query Match 14.9%; Score 41; DB 1; Length 260;
Best Local Similarity 41.78; Pred. No. 13;
Matches 10; Conservative 3; Mismatches 5; Indels 6; Gaps 1;

Oy 58 KGSVVD-----SVXGVFLTNQXD 75
DB 124 KGSVSDAVRASISIPGIFPQRLD 147

RESULT 9
RSMC_BUCAI STANDARD; PRT; 338 AA.
AC P57413;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribosomal RNA small subunit methyltransferase C (EC 2.1.1.52) (rRNA
DE (guanine-N(2)-methyltransferase) (16S rRNA m2G1207
DE methyltransferase).
DE RSMC OR B0328.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: SPECIFICALLY METHYLATES THE GUANOSINE IN POSITION 1207
CC OF 16S RNA IN THE 30S PARTICLE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-
CC homocysteine + rRNA containing N(2)-methylguanine.
CC -!- SIMILARITY: BELONGS TO THE METHYLTRANSFERASE SUPERFAMILY. RSMC
CC SUBFAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A POTENTIAL FRAMESHIFT

```

```

CC WAS CORRECTED IN POSITION 57.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001119; BAB13036.1; ALT_FRAME.
CC InterPro; IPR002052; N6_Mtase.
CC PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW rRNA processing; Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 338 AA; 39105 MW; E43516DDDD22FA014 CRC64;

Query Match 14.9%; Score 41; DB 1; Length 338;
Best Local Similarity 36.4%; Pred. No. 18;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Oy 57 WKSVDVSVXGVFLTNQNDXXS 78
DB 165 WKNLIKSLPGVFGHKIDSGS 186

RESULT 10
YHDP_BACSU STANDARD; PRT; 444 AA.
AC O07585;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yhdp.
CN YHDP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaere-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

```

CC -I- INDUCTION: Highest expression in darkness.  
CC -I- SIMILARITY: BELONGS TO THE AMT1/MEP/NRGA FAMILY OF AMMONIUM  
CC TRANSPORTERS (TC 2.49).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF118858; AAG11397.1; -  
DR InterPro; IPR001905; Ammonium\_transp.  
DR Pfam; PF00909; Ammonium\_transp; 1.  
DR TIGRFAMs; TIGR00836; ant; 1.  
KW PROSITE; PS01219; AMMONIUM\_TRANS; 1.  
KW Transport; Transmembrane; Multigene family.  
FT TRANSMEM 15 37 POTENTIAL.  
FT TRANSMEM 50 72 POTENTIAL.  
FT TRANSMEM 98 117 POTENTIAL.  
FT TRANSMEM 124 146 POTENTIAL.  
FT TRANSMEM 166 188 POTENTIAL.  
FT TRANSMEM 209 227 POTENTIAL.  
FT TRANSMEM 255 277 POTENTIAL.  
FT TRANSMEM 305 327 POTENTIAL.  
FT TRANSMEM 337 356 POTENTIAL.  
FT TRANSMEM 377 399 POTENTIAL.  
SQ SEQUENCE 460 AA; 49652 MW; 6D9CFC6C2C2395BA CRC64;  
  
Query Match 15.3%; Score 42; DB 1; Length 460;  
Best Local Similarity 42.9%; Pred. No. 17;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 57 WKGSVVDVSVXGVFLT 70  
t: |||||::|  
Db 5 WEASVTDTSINAIYL 18  
  
RESULT 7  
VG69\_BPPZA  
ID VG69\_BPPZA STANDARD; PRT; 108 AA.  
AC P08388;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Early protein GP16.9.  
GN 16.9.  
OS Bacteriophage PZA,  
OS Bacteriophage phi-15, and  
OS Bacteriophage phi-29.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
OC phi-29-like viruses.  
OX NCBI\_TaxID=10757, 10755, 10756;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=phage phi-15;  
RX MEDLINE=87031575; PubMed=3095189;  
RA Paces V., Vlcek C., Urbanek P., Hostomsky Z.,  
RA "Nucleotide sequence of the right early region of Bacillus subtilis  
RT phage PZA completes the 19366-bp sequence of PZA genome. Comparison  
RT with the homologous sequence of phage phi 29.";  
RL Gene 44:115-120(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=phage phi-15;  
RX MEDLINE=89232766; PubMed=2497055;  
RA Benes V., Arnold L., Smrt J., Paces V.,  
RA "Nucleotide sequence of the right early region of Bacillus phage phi  
RT 15 and comparison with related phages: reorganization of gene 17  
RT during evolution.";  
RL Gene 75:341-347(1989).  
RN [3]

RP SEQUENCE FROM N.A.  
RC SPECIES=phage phi-29;  
RX MEDLINE=86165872; PubMed=3007295;  
RA Garvey K.J., Yoshikawa H., Ito J.,  
RA "The complete sequence of the Bacillus phage phi 29 right early  
RT region.";  
RL Gene 40:301-309(1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M11813; AAA88498.1; -  
DR EMBL; M28830; AAA32330.1; -  
DR EMBL; M14430; AAA88350.1; -  
DR PIR; E29004; WRBP69.  
DR PIR; JS0193; WRBP13.  
DR PIR; JN0031; JN0031.  
KW Early protein.  
SQ SEQUENCE 108 AA; 12648 MW; 4FA9C051B60CD8F9 CRC64;  
  
Query Match 14.9%; Score 41; DB 1; Length 108;  
Best Local Similarity 35.0%; Pred. No. 5.3;  
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
  
Qy 52 RXFXWKGSVVDVSVXGVFLT 71  
t: |||||::|  
Db 59 RRFASWQGLLEGSNSVFWT 78  
  
RESULT 8  
YLBK\_BACSU  
ID YLBK\_BACSU STANDARD; PRT; 260 AA.  
AC O34731;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein ylbk.  
GN YLBK.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Bertero M., Presecan E., Glaser P., Richou A., Danchin A.;  
RA "Bacillus subtilis chromosomal region downstream nprE.";  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galletton N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

Query Match 15.6%; Score 43; DB 1; Length 474;  
Best Local Similarity 58.3%; Pred. No. 12;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 57 WKGSVVDSVXGV 68  
I:|||||I:|  
Db 89 WEGSVLDSEEGI 100

## RESULT 4

ID C1SY ASPNG STANDARD; PRT; 475 AA.  
AC P51044;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).  
GN CIT-1.  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WU-2223L;  
RA Oshida Y., Miyake K., Kanayama S., Kikimura K., Usami S.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H<sub>2</sub>O +  
CC oxaloacetate.  
CC -!- PATHWAY: Tricarboxylic acid cycle.  
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).  
CC -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS  
CC CAPABLE OF OXIDATIVE METABOLISM.  
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D63376; BAA09691.1; -;  
DR HSSP; P23007; 2CSC.  
DR InterPro; IPR002020; Citrate\_synt.  
DR Pfam; PF00285; citrate\_synt.1.  
DR PRINTS; PR00143; CITRATSYNTHASE.  
DR PROSITE; PS00480; CITRATE\_SYNTHASE; 1.  
KW Lyase; Tricarboxylic acid cycle; Mitochondrion; Transist peptide.  
FT TRANSIT 1 ? MITOCHONDRION (BY SIMILARITY).  
FT CHAIN ? 475 CITRATE SYNTHASE.  
FT ACT\_SITE 310 310 BY SIMILARITY.  
FT ACT\_SITE 356 356 BY SIMILARITY.  
FT ACT\_SITE 411 411 BY SIMILARITY.  
SQ SEQUENCE 475 AA; 5215 MW; F93525B3F3FCB3F CRC64;

Query Match 15.6%; Score 43; DB 1; Length 475;  
Best Local Similarity 58.3%; Pred. No. 12;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 57 WKGSVVDSVXGV 68  
I:|||||I:|  
Db 89 WEGSVLDSEEGI 100

## RESULT 5

ID SP1H\_HUMAN STANDARD; PRT; 232 AA.  
AC Q98655; O75650; Q9UJ0;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Spindlin homolog (Protein DXF34).  
GN DXF34.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97419273; PubMed=9271673;  
RA Laval S.H., Reed V., Blair H.J., Boyd Y.;  
RT "The structure of DXF34, a human X-linked sequence family with  
RT homology to a transcribed mouse Y-linked repeat."  
RL Mamm. Genome 8:689-691(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Heath P.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE SPIN / STSY FAMILY.  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y09858; CAA70988.1; -;  
DR EMBL; AL022157; CAA18148.1; -;  
DR EMBL; AL022157; CAA18149.1; -;  
DR InterPro; IPR003671; Spin\_Ssty.  
DR Pfam; PF02513; Spin-Ssty; 3.  
KW Developmental protein; Phosphorylation; Cell cycle.  
FT MOD\_RES 196 196 PHOSPHORYLATION (POTENTIAL).  
FT MOD\_RES 224 224 PHOSPHORYLATION (POTENTIAL).  
FT CONFLICT 20 20 C -> R (IN REF. 2).  
FT CONFLICT 208 208 T -> A (IN REF. 2; CAA18149).  
SQ SEQUENCE 232 AA; 26536 MW; 093E7E9F5340BF71 CRC64;

Query Match 15.3%; Score 42; DB 1; Length 232;  
Best Local Similarity 41.2%; Pred. No. 8.1;  
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 49 QGXRFXKXWKGVSVDV 65  
I:|||||I:|  
Db 34 EGDEPITQWKGTVLDQV 50

## RESULT 6

ID AT13\_LYCES STANDARD; PRT; 460 AA.  
AC Q9FVNO;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ammonium transporter 1, member 3 (Leamt1;3).  
GN AMT1.3.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Leaf;  
RX MEDLINE=20208060; PubMed=10743657;  
RA von Wiren N., Lauter F.-R., Ninnemann O., Gillissen B., Walch-Liu P.,  
RA Engels C., Jost W., Frommer W.B.;  
RT "Differential regulation of three functional ammonium transporter  
RT genes by nitrogen in root hairs and by light in leaves of tomato."  
RL Plant J. 21:167-175(2000).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- TISSUE SPECIFICITY: Leaves.

```
RL Nature 390:249-256(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY.
CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D84432; BAA12531.1; -.
DR EMBL; Z99116; CAB14406.1; -.
DR Subtilist; BG11695; yqba.
DR InterPro; IPR000644; CBS_domain.
DR InterPro; IPR005170; CorC_HlyC.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01595; DUF21; 1.
DR Pfam; PF03471; CorC_HlyC; 1.
KW - Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 16 36 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
SQ - SEQUENCE 442 AA; 49956 MW; E2EA41573549DA83 CRC64;

Query Match 17.1%; Score 47; DB 1; Length 442;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 62 VDSVXGFLTQNXSSXA 81
DB 396 VDTLGGWFLTQNDAPESA 415

RESULT 2
CISY_CANTR ID CISY_CANTR STANDARD; PRT; 467 AA.
AC P79024;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).
GN CIT.
OS Candida tropicalis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5482;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PK233.
RA Ueda M., Sanuki S., Kawachi H., Shimizu K., Atomi H., Tanaka A.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H(2)O +
CC oxaloacetate.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CC CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001565; BAA19410.1; -.
DR
```

```
DR HSSP; P23007; 2CSC.
DR InterPro; IPR002020; Citrate_synt.
DR Pfam; PF00285; citrate_synt; 1.
DR PRINTS; PR00143; CITRNTNTHASE.
DR PROSITE; PS00480; CITRATE-SYNTHASE; 1.
KW Lyase; Tricarboxylic acid cycle; MITOCHONDRION (POTENTIAL).
FT TRANSIT 1 467
FT CHAIN 1 467 BY SIMILARITY.
FT ACT_SITE 301 301 BY SIMILARITY.
FT ACT_SITE 347 347 BY SIMILARITY.
FT ACT_SITE 402 402 BY SIMILARITY.
SQ SEQUENCE 467 AA; 52004 MW; 4B194132C4198CA2 CRC64;

Query Match 15.6%; Score 43; DB 1; Length 467;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 57 WKGSVVDSVXGV 68
DB 79 WEGSVLDPIEG 90

RESULT 3
CISY_EMENI ID CISY_EMENI STANDARD; PRT; 474 AA.
AC O00098;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Citrate synthase, mitochondrial precursor (EC 4.1.3.7).
GN CIT.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=97306446; PubMed=9163747;
RA Park B.W., Han K.H., Lee C.Y., Lee C.H., Maeng P.J.;
RA "Cloning and characterization of the cita gene encoding the
RA mitochondrial citrate synthase of Aspergillus nidulans.";
RT Mitochondrial citrate synthase of Aspergillus nidulans.
RL Mol. Cells 7:290-295(1997).
CC -!- CATALYTIC ACTIVITY: Citrate + CoA = acetyl-CoA + H(2)O +
CC oxaloacetate.
CC -!- PATHWAY: Tricarboxylic acid cycle.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CC CAPABLE OF OXIDATIVE METABOLISM.
CC -!- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U89675; AAC49728.1; -.
DR HSSP; P23007; 2CSC.
DR InterPro; IPR002020; Citrate_synt.
DR Pfam; PF00285; citrate_synt; 1.
DR PRINTS; PR00143; CITRNTNTHASE.
DR PROSITE; PS00480; CITRATE-SYNTHASE; 1.
KW Lyase; Tricarboxylic acid cycle; MITOCHONDRION (POTENTIAL).
FT TRANSIT 1 35
FT CHAIN 36 474 BY SIMILARITY.
FT ACT_SITE 310 310 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT ACT_SITE 411 411 BY SIMILARITY.
SQ SEQUENCE 474 AA; 52223 MW; E2E86892ACB5398B CRC64;
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:14 ; Search time 3.23529 Seconds  
(without alignments)  
1153.796 Million cell updates/sec

Title: US-09-840-743-71

Perfect score: 275

Sequence: 1 KVXXDXTXXXXLXXXX.....TONXXSSXAMXXAXXFP 90

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	47	17.1	442	1 YQHB_BACSU	P54505 bacillus su
2	43	15.6	467	1 C1SY_CANTR	P79024 candida tro
3	43	15.6	474	1 C1SY_EMENI	O00098 emericella
4	43	15.6	475	1 C1SY_ASPEG	P51044 aspergillus
5	42	15.3	232	1 SP1H_HUMAN	Q99865 homo sapien
6	42	15.3	460	1 ATR13_LYCES	Q9fiv0 lycopersico
7	41	14.9	108	1 VG69_BPZPA	P08388 bacterioph
8	41	14.9	260	1 YLBK_BACSU	O34731 bacillus su
9	41	14.9	338	1 RSMC_BUCAI	P57413 buchera ap
10	41	14.9	444	1 YHDP_BACSU	O07585 bacillus su
11	41	14.9	1181	1 MYSC_DICDI	P42522 dictyosteli
12	41	14.9	1391	1 RPOB_MYCPN	P78013 mycoplasma
13	40	14.5	191	1 CLD7_RAT	Q92111 rattus norv
14	40	14.5	211	1 CLD7_MOUSE	Q92261 mus musculu
15	40	14.5	318	1 DEDD_HUMAN	Q75618 homo sapien
16	40	14.5	318	1 DEDD_MOUSE	Q75618 mus musculu
17	40	14.5	318	1 DEDD_RAT	Q922k0 rattus norv
18	40	14.5	469	1 C1SY_NEUCR	P34085 neurospora
19	40	14.5	796	1 YFS7_CABEL	Q10003 caenorhabdi
20	40	14.5	960	1 CHPL_SCHPO	Q10103 schizosacch
21	40	14.5	1503	1 MRP6_HUMAN	Q95255 homo sapien
22	40	14.5	1526	1 YK46_ANASP	Q8yrl1 anabaena sp
23	40	14.5	3911	1 AKA9_HUMAN	Q99996 h a-kinase
24	39.5	14.4	164	1 ATPF_STRPN	Q59952 streptococc
25	39	14.2	208	1 CT30_ARATH	Q91q17 arabidopsis
26	39	14.2	237	1 SP1N_HUMAN	Q9y657 homo sapien
27	39	14.2	240	1 SP1N_MOUSE	Q61142 mus musculu
28	39	14.2	268	1 COX3_SCHCO	P14058 schizophyil
29	39	14.2	270	1 YF53_AQUAE	O67438 aquifex ae
30	39	14.2	313	1 EBAG_STRPL	P04067 streptomyc
31	39	14.2	314	1 YG20_YEAST	P53257 saccharomyc
32	39	14.2	373	1 DP3B_MYCPU	Q98rk6 mycoplasma
33	39	14.2	389	1 UTL1_HUMAN	Q13336 homo sapien

RESULT 1

YQHB\_BACSU  
ID YQHB\_BACSU STANDARD: PRT: 442 AA.  
AC P54505;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein yqnb.

GN YQHB.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / JH642;

RX MEDLINE=97124195; PubMed=8969508;

RA Mizuno M., Masuda S., Takemaru K.I., Hosono S., Sato T., Takeuchi M.,

RA Kobayashi Y.;

RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of

RT the Bacillus subtilis genome containing the skin element and many

RL sporulation genes.";

RL Microbiology 142:3103-3111(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,

RA Piegsan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Totsato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

RT subtilis.";

**THIS PAGE BLANK (USPTO)**







New polynucleotide that control plant development comprising a sequence  
 having a specific homology to DEMETER domains A,B or C -  
 Disclosure; Page 91; 109pp; English.  
 The invention relates to an isolated polynucleotide sequence or their  
 complement encoding a polypeptide having a sequence at least 40%  
 identical to DMT (DEMIETER, previously known as ATROPOS (ATR)) Domain A, B  
 or C or their combinations. Also included are an expression  
 cassette comprising the polynucleotide or comprising a heterologous  
 polynucleotide under the control of a promoter at least 70% identical to  
 DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
 region of DMT, a host cell comprising an exogenous polynucleotide  
 encoding a DMT-like protein and a transgenic plant comprising a  
 polynucleotide encoding a DMT-like protein. The expression cassette is  
 useful for modulating transcription. The method comprises introducing  
 the cassette into a host cell preferably Agrobacterium by sexual  
 cross, and selecting a host cell with modulated transcription, where  
 the protein is capable of exhibiting at least one of the following  
 biological activities, which include enhanced expression of the  
 protein in a plant results in a delay in flowering time, introduction of  
 the protein into a cell results in modulation of methylation of  
 chromosomal DNA in the cell, reduction of expression of the protein in a  
 plant results in enhanced endosperm development and expressing of the  
 protein in an Arabidopsis leaf results in expression of the MEDA gene.  
 The polynucleotide is useful for detecting a nucleic acid in a sample.  
 DEMETER is related to 5-methylcytosine glycosylases and regulates  
 transcription of target genes by demethylation. The present  
 sequence represents a DMT-like protein.  
 Sequence 102 AA;

```

Query Match      15.0%; Score 88.5; DB 23; Length 102;
Best Local Similarity 28.3%; Pred. No. 1e-10;
Matches 28; Conservative 0; Mismatches 52; Indels 19; Gaps 3;

Qy 104 KXXXLRLTEHVVXXLPDXHXXLXXDXD-----XXYLLXIWPXXXXXX-----XXXXXX 152
      |  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 11 KNVSLRTEHCVELPDTPLHQLQWDTRPDDPGKYLAIWTPGTANSIQPESKCSQ 70

Qy 153 XXXXXXXXXXXXXXXXXXCCXXXCCXXXCCXXXCCXXXCCXXXCCXXXCCXXXCCXXXCCXXXCC 191
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 4 71 EEC-----GQLCNEHCFCSCNSFRANSQIVRGTL 101

```

RESULT 11	
AAU72760	
ID ✓	AAU72760 standard; Protein: 137 AA.
XX	
XX	AAU72760;
XX	
XX	26-FEB-2002 (first entry)
XX	
XX	Cotton DMT protein related sequence.
DE	
XX	
KW	Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;
KW	DNA demethylation; transgenic plant; transcription modulation;
KW	flowering time; endosperm development; MEDEA.
XX	
OS	Gossypium hirsutum.
XX	
XX	WO200180626-A1.
PN	
XX	
PD	01-NOV-2001.

XX  
PF 23-APR-2001; 2001WO-US13059.  
XX  
XX  
PR 21-APR-2000; 2000US-0553690.  
XX  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
XX  
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX

DR WPI; 2002-055307/07.  
DR N-PSDB; AAS96710.

XX New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A, B or C -  
PT  
XX  
PS Disclosure: Page 94: 109pp: English.

The invention relates to an isolated polynucleotide sequence or their complement encoding a polypeptide having a sequence at least 40% identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B or C or their combinations. Also included are an expression cassette comprising the polynucleotide or comprising a heterologous polynucleotide under the control of a promoter at least 70% identical to DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated region of DMT, a host cell comprising an exogenous polynucleotide encoding a DMT-like protein and a transgenic plant comprising a polynucleotide encoding a DMT-like protein. The expression cassette is useful for modulating transcription. The method comprises introducing the cassette into a host cell preferably *Agrobacterium* by sexual cross, and selecting a host cell with modulated transcription, where the protein is capable of exhibiting at least one of the following biological activities, which include enhanced expression of the protein in a plant results in a delay in flowering time, introduction of the protein into a cell results in modulation of methylation of chromosomal DNA in the cell, reduction of expression of the protein in a plant results in enhanced endosperm development and expressing of the protein in an *Arabidopsis* leaf results in expression of the MED2A gene. The polynucleotide is useful for detecting a nucleic acid in a sample. DEMETER is related to 5-methylcytosine glycosylases and regulates transcription of target genes by demethylation. The present sequence represents a DMT-like protein.

AA	Sequence	137 AA;
SQ		

Query Match	14.7%	Score 87;	DB 23;	Length 137;
Best Local Similarity	29.8%	Pred. No. 2.9e-10;		
Matches	28;	Conservative	0;	Mismatches 56;
				Indels 10;
				Gaps 3;

Qy	+	104 KXXXXLRTHEHVXXLPDXHXXLX-----XDXXXVLLXIWPXXXXXXXXXXXXXXXXCXX	157
Db	+	32 KNVSLRTEHYVVELDPKHPLLKOMKREPDSPYLIAIWTP-GETANSTOPPEOSC--88	

0v 158 XXXXXXXXXXXXXXXXXXXXXXXXXX 191

89 -GSOEPGRICNEKTCFACNSVBREANTEFTVBCTY 121

## RESULT 12

AAU72759  
ID AAU72759 standard; Protein: 164 AA.

AA  
AC AAI172759:XX  
DT 26-FEB-2002 (first entry)

DE: Corn DMT protein related sequence #1

Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEPEA.

Zea mays.

PN WO200180626-A1.

XX  
PD  
01-NOV-2001

23-APR-2001:

21-APR-2000: 2000US-0553690-XX  
PR

PA (REGC ) UNIV CALIFORNIA.

PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -  
XX  
XX Claim 5; SEQ ID NO 1471; 261pp + Sequence Listing; English.  
XX  
XX The invention relates to identifying target proteins  
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
CC aligning and comparing nucleic acid or amino acid sequences from plant  
CC with nucleic acid or amino acid sequences from non-plant organisms using  
CC suitable search parameters, where plant sequences having an E-value  
CC greater by a factor of 3 than the E-value of most similar non-plant  
CC sequences are selected. The polypeptides or nucleic acids encoding them  
CC are useful for identifying modulators. The identified modulators are  
CC useful as herbicides.  
XX  
XX Sequence 1309 AA;  
SQ

Query Match 18.4%; Score 108.5; DB 23; Length 1309;  
Best Local Similarity 25.0%; Pred. No. 7.6e-14;  
Matches 42; Conservative 0; Mismatches 117; Indels 9; Gaps 3;  
QY 104 KXXXXLATEHXXVXLPDXHXLL-----XXDXXXYYLLXIWXPXXXXXXXXXXXXXXXXXXCXX 157  
DB 1109 KIKKELTEHHVFELPDHHSILEGFERREAEIDIVYLIAIWTPL-GETVNSIQPPKQR-- 1165  
QY 158 XXXXXXXXCXXXXCXCCXXXEXXXXXXRGTXLXXXXXXXXXXXXXXXXXXXXXXXFPAD 217  
DB 1166 ALPENNTLCNENKCFQCNTREEESQTVRCTILIPCTAMRGGFPLNGTYFQTNEVPAD 1225  
QY 218 HXXXXXPPIXXXXXXMLXPRXXXXGXGXXXXXXLXXXXXXIXXFXPG 265  
DB 1226 HDSSINPIDVPTLIWDLKRRVAYLGSSVSICKGLSVEAIKNFQEG 1273

RESULT 9  
AAU72745 ID AAU72745 standard; Protein; 177 AA.  
XX AAU72745;  
XX AC  
XX DT 26-FEB-2002 (first entry)  
XX XX  
DE Corn DMT1 protein.  
XX  
KW Demeter; DMT; Atropis; ATR: 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
OS Zea mays.  
XX  
XX WO200180626-A1.  
XX PN  
XX PD 01-NOV-2001.  
XX XX  
XX PF 23-APR-2001; 2001WO-US13059.  
XX XX  
XX PR 21-APR-2000; 2000US-0553690.  
XX XX  
XX (REGC ) UNIV CALIFORNIA.  
XX PA  
XX PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX DR WPI: 2002-055307/07.  
DR N-PSDB; AAS96696.  
XX

New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A,B or C -  
XX  
XX Disclosure; Page 84; 109pp; English.  
XX

The invention relates to an isolated polynucleotide sequence or their  
CC complement encoding a polypeptide having a sequence at least 40%  
CC

CC protein in a plant results in a delay in flowering time, introduction of  
CC the protein into a cell results in modulation of methylation of  
CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expressing of the  
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.  
XX  
SQ Sequence 1114 AA;  
  
Query Match 22.8%; Score 134.5; DB 23; Length 1114;  
Best Local Similarity 25.8%; Pred. No. 1.8e-19;  
Matches 49; Conservative 0; Mismatches 108; Indels 33; Gaps 4;  
  
QY 104 KXXXXLRTTEHXXVXXLPDXHXLL-----XXDXXXVLLXIWXPPXXXXXXXCCXX 157  
DB 899 KYNRLRTEHVVVLPDNLHELHDFERRKLDPSYLLAIWOP-GETSSFPVPPKKC-- 955  
QY 158 XXXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 217  
DB 956 --SSDGSKLCKIKNCYSYCWTFIREQNSNIFRGITL-----VFAD 991  
QY 218 HXXXXXXPIXXXXXXLLXXRXXXXXXGXXXXXXIXXXLXXXXXXGXXCXXRFXXXXX 277  
DB 992 HETSLNPIVFRRELCKGLEKRALYCGSTVTSIFKLLDTRRIELCFWTGFLCLRAFDKOR 1051  
  
QY 278 XPXXLXXLX 287  
DB 1052 DPXELVRLH 1061  
  
RESULT 7  
AAU72757  
ID AAU72757 standard; Protein; 210 AA.  
XX  
AC AAU72757;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Tomato DMT protein related sequence #1.  
XX  
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX  
OS Lycopodium esculentum.  
XX  
PN WO200180626-A1.  
XX  
PD 01-NOV-2001.  
XX  
PF 23-APR-2001; 2001WO-US13059.  
XX  
PR 21-APR-2000; 2000US-0553690.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX  
DR WPI; 2002-055307/07.  
DR N-P5DB; AAS96707.  
XX  
PT New polynucleotide that control plant development comprising a sequence  
PT having a specific homology to DEMETER domains A,B or C  
XX  
PS Disclosure; Page 92; 109pp; English.  
XX  
CC The invention relates to an isolated polynucleotide sequence or their  
CC complement encoding a polypeptide having a sequence at least 40%  
CC identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
CC or C or their combinations. Also included are an expression

CC cassette comprising the polynucleotide or comprising a heterologous  
CC polynucleotide under the control of a promoter at least 70% identical to  
CC DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
CC region of DMT, a host cell comprising an exogenous polynucleotide  
CC encoding a DMT-like protein and a transgenic plant comprising a  
CC polynucleotide encoding a DMT-like protein. The expression cassette is  
CC useful for modulating transcription. The method comprises introducing  
CC the cassette into a host cell preferably Agrobacterium by sexual  
CC cross, and selecting a host cell with modulated transcription, where  
CC the protein is capable of exhibiting at least one of the following  
CC biological activities, which include enhanced expression of the  
CC protein in a plant results in a delay in flowering time, introduction of  
CC the protein into a cell results in modulation of methylation of  
CC chromosomal DNA in the cell, reduction of expression of the protein in a  
CC plant results in enhanced endosperm development and expressing of the  
CC protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
CC The polynucleotide is useful for detecting a nucleic acid in a sample.  
CC DEMETER is related to 5-methylcytosine glycosylases and regulates  
CC transcription of target genes by demethylation. The present  
CC sequence represents a DMT-like protein.  
XX  
SQ Sequence 210 AA;  
  
Query Match 22.4%; Score 132; DB 23; Length 210;  
Best Local Similarity 24.9%; Pred. No. 1.1e-19;  
Matches 47; Conservative 0; Mismatches 124; Indels 18; Gaps 3;  
  
QY 104 KXXXXLRTTEHXXVXXLPDXHXLLX-----DXXXVLLXIWXPPXXXXXXX 153  
DB 28 KNVSRRLRTEHVSVELPDSHPLEGEKREPDPDGKYLAIWTFGETANSIQPDRCSAQ 87  
QY 154 XCXXXXXXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXXCCXX 213  
DB 88 DC-----GOLCNEECFSCNSFREANSQIVRGITILIPCTAMRGSPFLNGTVQVNE 139  
QY 214 XFADHXXXXXXPIXXXXXXLLXXRXXXXXXGXXXXXXIXXXLXXXXXXGXXCXXRFX 273  
DB 140 VFADHESLNPIVPRSLIWNLDRTVHFGTSVTSIFKGLATPEIQOCFWRGFCVRSFE 199  
  
QY 274 XXXXXPPXXL 282  
DB 200 RSTRAPRPL 208  
  
RESULT 8  
ABB92260  
ID ABB92260 standard; Protein; 1309 AA.  
XX  
AC ABB92260;  
XX  
DT 31-MAY-2002 (first entry)  
XX  
DE Herbicidally active polypeptide SEQ ID NO 1471.  
XX  
KW Herbicidal; plant; agriculture; herbicide.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200210210-A2.  
PD 07-FEB-2002.  
XX  
PF 28-AUG-2001; 2001WO-EP09892.  
XX  
PR 28-AUG-2001; 2001WO-EP09892.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Tietjen K, Weidler M;  
XX  
DR WPI; 2002-269010/31.  
XX  
PT Identifying plant target proteins for herbicidally active compounds,



KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX  
OS Oryza sativa.  
XX WO200180626-A1.  
XX 01-NOV-2001.  
XX 23-APR-2001; 2001WO-US13059.  
XX 21-APR-2000; 2000US-0553690.  
XX (RECC ) UNIV CALIFORNIA.  
XX Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX WPI; 2002-055307/07.  
XX N-PSDB; AAS96695.  
XX New polynucleotide that control plant development comprising a sequence  
XX having a specific homology to DEMETER domains A,B or C  
XX  
XX Disclosure; Page 79-80; 109pp; English.  
XX The invention relates to an isolated polynucleotide sequence or their  
XX complement encoding a polypeptide having a sequence at least 40%  
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
XX or C or their combinations. Also included are an expression  
XX cassette comprising the polynucleotide or comprising a heterologous  
XX polynucleotide under the control of a promoter at least 70% identical to  
XX DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
XX region of DMT, a host cell comprising an exogenous polynucleotide  
XX encoding a DMT-like protein and a transgenic plant comprising a  
XX polynucleotide encoding a DMT-like protein. The expression cassette is  
XX useful for modulating transcription. The method comprises introducing  
XX the cassette into a host cell preferably Agrobacterium by sexual  
XX cross, and selecting a host cell with modulated transcription, where  
XX the protein is capable of exhibiting at least one of the following  
XX biological activities, which include enhanced expression of the  
XX protein in a plant results in a delay in flowering time, introduction of  
XX the protein into a cell results in modulation of methylation of  
XX chromosomal DNA in the cell, reduction of expression of the protein in a  
XX plant results in enhanced endosperm development and expression of the  
XX protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
XX The polynucleotide is useful for detecting a nucleic acid in a sample.  
XX DEMETER is related to 5-methylcytosine glycosylases and regulates  
XX transcription of target genes by demethylation. The present  
XX sequence represents a DMT-like protein.  
XX  
SQ Sequence 1952 AA;  
  
Query Match 24.6%; Score 145; DB 23; Length 1952;  
Best Local Similarity 26.2%; Pred. No. 1.8e-21;  
Matches 51; Conservative 0; Mismatches 134; Indels 10; Gaps 3;  
  
QY 104 KXXXLRLRTHVXXLPDXHXXLXX-----XDXXXYLXIXWXPXXXXXXXCX 157  
DB 1745 KXVSRRLRTHVQVYELPSHPLLEGFNQREPDPCPYLLSIWTP-GETAQSTDAKPSVC-- 1801  
QY 158 XXXXXXXXXXCCXXCCXXXEXXXXXXGTXLXXXXXXXXXXXXXXXXXXXXFAD 217  
DB 1802 -NSQNGELCASNTCFSCNSTREAQAQVGTLLIPCTAMRGSPPLNGTVFQVNEVFAD 1860  
QY 218 HXXXXXPXXXXXXXLXRXXXXXXXXGXIXXXXXXXXIXXFXGXGXCCXRXFXXXXX 277  
DB 1861 HDSSRNPDVPRSWLWNLPRRTVFGTISPTIFKGLTTEIQCWFRCVCGRFDRTSR 1920  
QY 278 XPXXLXXLHXXXXX 292  
DB 1921 APRPLYARLHFPAK 1935

RESULT 4  
AAU72756  
ID AAU72756 standard; Protein; 217 AA.  
XX  
AC AAU72756;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Alfalfa DMT protein related sequence.  
XX  
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX  
OS Medicago sativa.  
XX WO200180626-A1.  
XX  
PD 01-NOV-2001.  
XX 23-APR-2001; 2001WO-US13059.  
XX 21-APR-2000; 2000US-0553690.  
XX (RECC ) UNIV CALIFORNIA.  
XX Fischer RL, Choi Y, Hannon M, Okamuro JK, Tatarinova TV;  
XX WPI; 2002-055307/07.  
XX N-PSDB; AAS96706.  
XX New polynucleotide that control plant development comprising a sequence  
XX having a specific homology to DEMETER domains A,B or C  
XX  
XX Disclosure; Page 92; 109pp; English.  
XX The invention relates to an isolated polynucleotide sequence or their  
XX complement encoding a polypeptide having a sequence at least 40%  
XX identical to DMT (DEMETER, previously known as ATROPOS (ATR)) Domain A, B  
XX or C or their combinations. Also included are an expression  
XX cassette comprising the polynucleotide or comprising a heterologous  
XX polynucleotide under the control of a promoter at least 70% identical to  
XX DMT 5' flanking sequence, DMT 3' flanking sequence or an 5' untranslated  
XX region of DMT, a host cell comprising an exogenous polynucleotide  
XX encoding a DMT-like protein and a transgenic plant comprising a  
XX polynucleotide encoding a DMT-like protein. The expression cassette is  
XX useful for modulating transcription. The method comprises introducing  
XX the cassette into a host cell preferably Agrobacterium by sexual  
XX cross, and selecting a host cell with modulated transcription, where  
XX the protein is capable of exhibiting at least one of the following  
XX biological activities, which include enhanced expression of the  
XX protein in a plant results in a delay in flowering time, introduction of  
XX the protein into a cell results in modulation of methylation of  
XX chromosomal DNA in the cell, reduction of expression of the protein in a  
XX plant results in enhanced endosperm development and expression of the  
XX protein in an Arabidopsis leaf results in expression of the MEDEA gene.  
XX The polynucleotide is useful for detecting a nucleic acid in a sample.  
XX DEMETER is related to 5-methylcytosine glycosylases and regulates  
XX transcription of target genes by demethylation. The present  
XX sequence represents a DMT-like protein.  
XX  
SQ Sequence 217 AA;  
  
Query Match 24.2%; Score 143; DB 23; Length 217;  
Best Local Similarity 25.3%; Pred. No. 4.9e-22;  
Matches 50; Conservative 0; Mismatches 130; Indels 18; Gaps 3;  
  
QY 104 KXXXLRLRTHVXXLPDXHXXLXX-----XDXXXYLXIXWXPXXXXXXX 153  
DB 28 KXVSRRLRTHVQVYELPSHPLLEGWKEKREPDPCGYLLAIWTPGTANSIQPPDRCSAQ 87  
QY 154 XXXXXXXXXXCCXXCCXXXEXXXXXXGTXLXXXXXXXXXXXXXXXXXXXX 213



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:08 ; Search time 45.3268 Seconds  
(without alignments)  
858.415 Million cell updates/sec

Title: US-09-840-743-73

Perfect score: 590

Sequence: 1 PXEXPPXXXXXEXXXXXX.....XXXXXXPLXXLHXXXXXK 292

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Genes2\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	24.7	1729	23 AAU72736	Arabidopsis Demete
2	145	24.6	1413	23 AAU72737	Arabidopsis DMT2 (
3	145	24.5	1952	23 AAU72743	Rice DMT1 protein.
4	143	24.2	217	23 AAU72756	Alfalfa DMT protei
5	137.5	23.3	1332	23 AAU72739	Arabidopsis DMT3 (
6	134.5	22.8	1114	23 AAU72741	Arabidopsis DMT4 (
7	132	22.4	210	23 AAU72757	Tomato DMT protein
8	108.5	18.4	1309	23 ABB92260	Herbicidally activ
9	100	16.9	177	23 AAU72745	Corn DMT1 protein.
10	88.5	15.0	102	23 AAU72755	Soybean DMT4 prote

11	87	14.7	137	23 AAU72760	Cotton DMT protein
12	87	14.7	164	23 AAU72759	Corn DMT protein r
13	85	14.4	179	23 AAU72764	Corn DMT protein r
14	79	13.4	209	23 AAU72754	Soybean DMT3 prote
15	77	13.1	108	23 AAU72750	Wheat DMT3 protein
16	77	13.1	140	23 AAU72766	Corn DMT protein r
17	76	12.9	138	23 AAU72767	Corn DMT protein r
18	74	12.5	150	23 AAU72765	Corn DMT protein r
19	72	12.2	89	23 AAU72768	Corn DMT protein r
20	66	11.2	93	23 AAU72769	Corn DMT protein r
21	59	10.0	183	23 AAU72748	Wheat DMT1 protein
22	54	9.2	71	23 AAU72770	Corn DMT protein r
23	41	6.9	1548	22 ABB62738	Drosophila melanog
24	41	6.9	1548	22 AAB72443	UGT. Drosophila
25	41	6.9	1548	23 AAM47583	Drosophila cell cy
26	37	6.3	229	21 AAU90301	Human peptidase, H
27	37	6.3	229	22 AAU12260	Human PRO4389 poly
28	37	6.3	233	22 AAM92597	Human digestive sy
29	37	6.3	233	22 AAM38606	Human colorectal c
30	37	6.3	3503	22 ABB66499	Drosophila melanog
31	36	6.1	85	22 AAM88162	Human immune/haema
32	36	6.1	144	21 AAG35357	Zea mays protein f
33	36	6.1	204	22 AAU50869	Propionibacterium
34	36	6.1	662	21 AAU95685	Cosmid cHRIM5 enco
35	35	5.9	75	22 AAU48812	Propionibacterium
36	35	5.9	81	22 AAU53845	Propionibacterium
37	35	5.9	606	20 AAU43219	E. thsiopathiae s
38	34.5	5.8	1289	22 ABB70840	Drosophila melanog
39	34	5.8	49	21 AAG27372	Arabidopsis thalia
40	34	5.8	83	22 AAU22769	Human prostate can
41	34	5.8	83	22 AAM94749	Human reproductive
42	34	5.8	109	21 AAG06711	Arabidopsis thalia
43	34	5.8	112	21 AAG06710	Arabidopsis thalia
44	34	5.8	152	23 AAU10493	Human HOMO tyrosin
45	34	5.8	154	22 AAB60106	Human transport pr

ALIGNMENTS

RESULT 1  
AAU72736  
ID AAU72736 standard; Protein: 1729 AA.  
XX AAU72736;  
AC AC  
XX XX  
DT 26-FEB-2002 (first entry)  
XX XX  
DE Arabidopsis Demeter protein.  
XX XX  
KW Demeter; DMT; Atropos; ATR; 5-methylcytosine glycosylase;  
KW DNA demethylation; transgenic plant; transcription modulation;  
KW flowering time; endosperm development; MEDEA.  
XX XX  
OS Arabidopsis Thaliana.  
XX OS  
PN WO200180626-A1.  
XX XX  
PD 01-NOV-2001.  
XX XX  
PF 23-APR-2001; 2001WO-US13059.  
XX XX  
PR 21-APR-2000; 2000US-0553690.  
XX XX  
PA (REGC ) UNIV CALIFORNIA.  
XX XX  
PI Fischer RL, Choi Y, Hannon M, Okumuro JK, Tatarinova TV;  
XX XX  
DR WPI; 2002-055307/07.  
XX XX  
DR N-PSDB; AAS96686, AAS96689.  
XX XX  
PT New polynucleotide that control plant development comprising a sequence having a specific homology to DEMETER domains A, B or C -



**THIS PAGE BLANK (USPTO)**

; REFERENCE/DOCKET NUMBER: DX0335K1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 578 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-110-683-2

Query Match 5.43; Score 32; DB 1; Length 578;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 109 LRTEHXVXXLPDXH 122  
|:|:| | | |  
Db 328 LQTEEPQFLPDPH 341

Search completed: November 8, 2002, 19:29:22  
Job time : 22.6993 secs

QY 133 YLLXIWP 140  
| | | : | |  
Db 976 YLLSLWP 983

RESULT 13  
US-08-424-788-6  
; Sequence 6, Application US/08424788  
; Patent No. 5716804  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Kevin W.  
; APPLICANT: Wei, Sherry  
; APPLICANT: Ho, Alice Suk-Yue  
; TITLE OF INVENTION: MAMMALIAN INTERLEUKIN-10 (IL-10)  
; TITLE OF INVENTION: SUPER-ACTIVATING RECEPTORS; AND VARIANTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,788  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0501  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 557 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-424-788-6

Query Match 5.4%; Score 32; DB 1; Length 557;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXXVXXLPDXH 122  
| : | | |  
Db 307 LQTEPQFLLPDPH 320

RESULT 14  
US-08-424-788-5  
; Sequence 5, Application US/08424788  
; Patent No. 5716804  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Kevin W.  
; APPLICANT: Wei, Sherry  
; APPLICANT: Ho, Alice Suk-Yue  
; TITLE OF INVENTION: MAMMALIAN INTERLEUKIN-10 (IL-10)  
; TITLE OF INVENTION: SUPER-ACTIVATING RECEPTORS; AND VARIANTS  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto

; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,788  
; FILING DATE: 19-APR-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0501  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-852-9196  
; TELEFAX: 415-496-1200  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 578 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-424-788-5

Query Match 5.4%; Score 32; DB 1; Length 578;  
Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXXVXXLPDXH 122  
| : | | |  
Db 328 LQTEPQFLLPDPH 341

RESULT 15  
US-08-110-683-2  
; Sequence 2, Application US/08110683  
; Patent No. 5789192  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Kevin W.  
; APPLICANT: Liu, Ying  
; APPLICANT: Ho, Alice Suk-Yue  
; APPLICANT: Hsu, Di-Hwei  
; APPLICANT: Bazan, J. Fernando  
; APPLICANT: Tan, Jimmy C.  
; APPLICANT: Chou, Chuan-Chu  
; TITLE OF INVENTION: Mammalian Receptors for Interleukin-10  
; TITLE OF INVENTION: (IL-10)  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/110,683  
; FILING DATE: 23-AUG-1993  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/011,066  
; FILING DATE: 29-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090

;; APPLICANT: Williams, Lewis T.  
;; APPLICANT: Molz, Lisa  
;; APPLICANT: Chen, Yen-Wen  
;; TITLE OF INVENTION: No. 594864el PI 3-Kinase Polypeptides  
;; NUMBER OF SEQUENCES: 32  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, 8th Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/609,049A  
;; FILING DATE: 29-FEB-1996  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dow, Karen B.  
;; REGISTRATION NUMBER: 29,684  
;; REFERENCE/DOCKET NUMBER: 2307K-0637000S  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-326-2400  
;; TELEFAX: 415-326-2422  
;; INFORMATION FOR SEQ ID NO: 28:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1876 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-609-049A-28

Query Match 5.6%; Score 33; DB 2; Length 1876;  
Best Local Similarity 62.5%; Pred. No. 3.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 133 YLLXIWPX 140  
||| :| |  
Db - 976 YLLSLWPP 983  
RESULT 11  
US-09-170-996-12  
;; Sequence 12, Application US/09170996  
;; Patent No. 6291220  
;; GENERAL INFORMATION:  
;; APPLICANT: Williams, Lewis T.  
;; APPLICANT: Molz, Lisa  
;; APPLICANT: Chen, Yen-Wen  
;; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides  
;; NUMBER OF SEQUENCES: 32  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, 8th Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/170,996  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/609,049

;; FILING DATE: 29-FEB-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dow, Karen B.  
;; REGISTRATION NUMBER: 29,684  
;; REFERENCE/DOCKET NUMBER: 2307K-0637000S  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-326-2400  
;; TELEFAX: 415-326-2422  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1876 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-170-996-12  
Query Match 5.6%; Score 33; DB 4; Length 1876;  
Best Local Similarity 62.5%; Pred. No. 3.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
Oy 133 YLLXIWPX 140  
||| :| |  
Db 976 YLLSLWPP 983  
RESULT 12  
US-09-170-996-28  
;; Sequence 28, Application US/09170996  
;; Patent No. 6291220  
;; GENERAL INFORMATION:  
;; APPLICANT: Williams, Lewis T.  
;; APPLICANT: Molz, Lisa  
;; APPLICANT: Chen, Yen-Wen  
;; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides  
;; NUMBER OF SEQUENCES: 32  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Townsend and Townsend and Crew LLP  
;; STREET: Two Embarcadero Center, 8th Floor  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-3834  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/170,996  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/609,049  
;; FILING DATE: 29-FEB-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dow, Karen B.  
;; REGISTRATION NUMBER: 29,684  
;; REFERENCE/DOCKET NUMBER: 2307K-0637000S  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-326-2400  
;; TELEFAX: 415-326-2422  
;; INFORMATION FOR SEQ ID NO: 28:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1876 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-170-996-28  
Query Match 5.6%; Score 33; DB 4; Length 1876;  
Best Local Similarity 62.5%; Pred. No. 3.3e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 60 DTRYLADVWP 71

1 11 : 1 1

RESULT 7

US-08-899-514-2

Sequence 2, Application US/08899514

Patent No. 5910581

GENERAL INFORMATION:

APPLICANT: HABUCHI, OSAMI

APPLICANT: FUKUTA, MASAKAZU

TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN

TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING

TITLE OF INVENTION: FOR THE POLYPEPTIDE

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP

STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR

CITY: NEWPORT BEACH

STATE: CALIFORNIA

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,514

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: DANIEL E. ALTMAN

REGISTRATION NUMBER: 34,115

REFERENCE/DOCKET NUMBER: TOYAW21.001AUS

TELECOMMUNICATION INFORMATION:

TELEPHONE: 714 760 0404

TELEFAX: 714 760 9502

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 479

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-899-514-2

Query Match

Best Local Similarity 45.5%; Score 33; DB 2; Length 479;

Matches 5; Conservative 2; Mismatches 2; Indels 4; Gaps 0;

QY 112 EHVXVXLDPXH 122

11 : 11 : 1

Db 204 EHFITPLPEDH 214

RESULT 8

US-09-172-977-4

Sequence 4, Application US/09172977

Patent No. 5989863

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.

APPLICANT: Yue, Henry

TITLE OF INVENTION: HUMAN ANKYRIN FAMILY PROTEIN

FILE REFERENCE: PF-0615 US

CURRENT APPLICATION NUMBER: US/09/172,977

CURRENT FILING DATE: 1998-10-14

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PERL Program

10

SEQ ID NO 4

LENGTH: 1839

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE: -

OTHER INFORMATION: g29491

US-09-172-977-4

Query Match

Best Local Similarity 60.0%; Score 33; DB 2; Length 1839;

Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 129 DXXYYLLXIW 138

1 111 11

Db 1493 DQSYLLKIW 1502

RESULT 9

US-08-609-049A-12

Sequence 12, Application US/08609049A

Patent No. 5948664

GENERAL INFORMATION:

APPLICANT: Williams, Lewis T.

APPLICANT: Molz, Lisa

APPLICANT: Chen, Yen-Wen

TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/609,049A

FILING DATE: 29-FEB-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Dow, Karen B.

REGISTRATION NUMBER: 29,684

REFERENCE/DOCKET NUMBER: 2307K-0637000US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-326-2400

TELEFAX: 415-326-2422

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1876 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-609-049A-12

Query Match

Best Local Similarity 62.5%; Score 33; DB 2; Length 1876;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 133 YLLXIXWP 140

111 : 1

Db 976 YLLSLWPP 983

RESULT 10

US-08-609-049A-28

Sequence 28, Application US/08609049A

Patent No. 5948664

GENERAL INFORMATION:

US-09-036-987A-22

Query Match 5.8%; Score 34; DB 4; Length 278;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
| | | | | | | | | |  
DB 159 LNTQQVAILPTSH 172

RESULT 4

US-09-370-700-22  
Sequence 22, Application US/09370700  
Patent No. 6274350

GENERAL INFORMATION:  
APPLICANT: Baltz, Richard H  
APPLICANT: Broughton, Mary C  
APPLICANT: Crawford, Kathryn P  
APPLICANT: Madduri, Krishnamurthy  
APPLICANT: Treadway, Patti J  
APPLICANT: Turner, Jan R  
APPLICANT: Waldron, Clive  
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
FILE REFERENCE: 50489 DIV1  
CURRENT APPLICATION NUMBER: US/09/370,700  
CURRENT FILING DATE: 1999-08-09  
EARLIER APPLICATION NUMBER: US 09/36987  
EARLIER FILING DATE: 1998-03-09  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 22  
LENGTH: 278  
TYPE: PRT  
ORGANISM: Saccharopolyspora spinosa

US-09-370-700-22

Query Match 5.8%; Score 34; DB 4; Length 278;  
Best Local Similarity 50.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
| | | | | | | | | |  
DB 159 LNTQQVAILPTSH 172

RESULT 5

US-09-071-035-76  
Sequence 76, Application US/09071035  
Patent No. 6448043

GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369p2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 76:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 100 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-76

Query Match 5.6%; Score 33; DB 4; Length 100;  
Best Local Similarity 41.7%; Pred. No. 17;  
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 129 DXXXYLXIXWP 140  
| | | | | | | | | |  
DB 41 DTNRYLADVWVP 52

RESULT 6

US-09-071-035-74  
Sequence 74, Application US/09071035  
Patent No. 6448043

GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369p2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 74:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 119 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-74

Query Match 5.6%; Score 33; DB 4; Length 119;  
Best Local Similarity 41.7%; Pred. No. 20;  
Matches 5; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 129 DXXXYLXIXWP 140

COUNTRY: USA  
ZIP: 60603-2828  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" diskette  
COMPUTER: Compaq Prolinea 5100  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685.466C  
FILING DATE: 24-JUL-96  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/001.381  
FILING DATE: 24-JUL-95  
ATTORNEY/AGENT INFORMATION:  
NAME: Daniel J. Hulseberg  
REGISTRATION NUMBER: 36,554  
REFERENCE/DOCKET NUMBER: 96303755  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 701-8606  
TELEFAX: (312) 701-7711  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 192 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified amino acid, formylmethionine  
LOCATION: The first amino acid in the sequence,  
LOCATION: described as methionine, is actually known as  
LOCATION: formylmethionine ("fMet"). Polypeptide chains in  
LOCATION: bacteria often start with formylmethionine.  
PUBLICATION INFORMATION:  
AUTHORS: L. de Boer, W. Harder, L. Dijkhuizen  
TITLE: Phenylalanine and tyrosine metabolism in the  
TITLE: facultative methyloleotroph No. 6015705ardia sp. 239  
JOURNAL: Arch. Microbiol.  
VOLUME: 149  
PAGES: 459-465  
DATE: 1988  
PUBLICATION INFORMATION:  
AUTHORS: J. Cairns, J. Overbaugh, S. Miller  
TITLE:  
JOURNAL: Nature  
VOLUME: 335  
PAGES: 142-145  
DATE: 1988  
PUBLICATION INFORMATION:  
AUTHORS: G.J.W. Euverink, G.I. Hessels, J.W.  
AUTHORS: Vrijbloed, J.R. Coggins, L. Dijkhuizen  
TITLE: Purification and characterization of a dual  
TITLE: function 3-hydroquinone dehydratase from Amycolatopsis  
TITLE: methanolica  
JOURNAL: J. Gen. Microbiology  
VOLUME: 138  
PAGES: 2449-2457  
DATE: 1992  
PUBLICATION INFORMATION:  
AUTHORS: G.J.W. Euverink, D.J. Wolters, L.  
AUTHORS: Dijkhuizen  
TITLE: Prephenate dehydratase of the actinomycete  
TITLE: Amycolatopsis methanolica: purification and  
TITLE: characterization of the wild-type and deregulated mutant  
TITLE: enzymes  
JOURNAL: Biochem. J.  
VOLUME: 308  
PAGES: 313-320  
DATE: 1995  
PUBLICATION INFORMATION:  
AUTHORS: P. Moretti, G. Hintermann, R. H tter  
TITLE: Isolation and characterization of an  
TITLE: extrachromosomal element from No. 6015705ardia mediterranei  
JOURNAL: Plasmid

VOLUME: 14  
PAGES: 126-133  
DATE: 1985  
PUBLICATION INFORMATION:  
AUTHORS: J. Sambrook, E.F. Fritsch, T. Maniatis  
TITLE: Molecular cloning: a laboratory manual  
JOURNAL: Cold Spring Harbor Laboratory Press  
DATE: 1989  
PUBLICATION INFORMATION:  
AUTHORS: J.W. Vrijbloed, J. Madon, L. Dijkhuizen  
TITLE: A plasmid from the methyloleotroph  
TITLE: actinomycete Amycolatopsis methanolica capable of site-  
TITLE: specific integration  
JOURNAL: J. Bacteriol.  
VOLUME: 176  
PAGES: 7087-7090  
DATE: 1994  
US-08-685-466C-4

Query Match 5.8%; Score 34; DB 3; Length 192;  
Best Local Similarity 24.1%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

QY 110 RTEHVXXLPDXHXXLXXXXXXLLXIW 138  
DB 47 RFEHTVWPPVMAIVFNDTADHVLMAW 75

RESULT 3  
US-09-036-987A-22  
; Sequence 22, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patti J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow AgroSciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036.987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 278 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:15 ; Search time 16.6993 Seconds  
(without alignments)  
514.481 Million cell updates/sec

Title: US-09-840-743-73

Perfect score: 590

Sequence: 1 PXXEXPPXXXXXXEXXXXXX.....XXXXXXXXXXLHXXXXXK 292

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	6.9	1548	US-09-376-330-15	Sequence 15, Appl
2	34	5.8	192	US-08-685-466C-4	Sequence 4, Appl
3	34	5.8	278	US-09-036-987A-22	Sequence 22, Appl
4	34	5.8	278	US-09-370-700-22	Sequence 22, Appl
5	33	5.6	100	US-09-071-035-76	Sequence 76, Appl
6	33	5.6	119	US-09-071-035-74	Sequence 74, Appl
7	33	5.6	479	US-08-899-514-2	Sequence 2, Appl
8	33	5.6	1839	US-09-172-977-4	Sequence 4, Appl
9	33	5.6	1876	US-08-609-049A-12	Sequence 12, Appl
10	33	5.6	1876	US-08-609-049A-28	Sequence 28, Appl
11	33	5.6	1876	US-09-170-996-12	Sequence 12, Appl
12	33	5.6	1876	US-09-170-996-28	Sequence 28, Appl
13	32	5.4	557	US-08-424-788-6	Sequence 6, Appl
14	32	5.4	578	US-08-424-788-5	Sequence 5, Appl
15	32	5.4	578	US-08-110-683-2	Sequence 2, Appl
16	32	5.4	578	US-08-683-743-2	Sequence 2, Appl
17	32	5.4	578	US-08-477-166-2	Sequence 2, Appl
18	32	5.4	578	US-08-472-097-2	Sequence 2, Appl
19	32	5.4	578	US-09-439-672-2	Sequence 2, Appl
20	32	5.4	578	PCT-US93-11638-2	Sequence 2, Appl
21	31	5.3	206	US-09-171-611-13	Sequence 13, Appl
22	31	5.3	223	US-08-928-612-3	Sequence 3, Appl
23	31	5.3	323	US-09-199-637A-215	Sequence 215, App
24	31	5.3	434	US-08-815-718-3	Sequence 3, Appl
25	31	5.3	457	US-08-206-006-2	Sequence 2, Appl
26	31	5.3	518	US-09-134-001C-5040	Sequence 5040, Ap
27	31	5.3	559	US-08-884-072-6	Sequence 6, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-376-330-15  
; Sequence 15, Application US/09376330  
; Patent No. 6399321  
; GENERAL INFORMATION:  
; APPLICANT: Tessier, Daniel C.  
; APPLICANT: Dignard, Daniel  
; APPLICANT: Bergeron, John J.M.  
; APPLICANT: Thomas, David Y.  
; TITLE OF INVENTION: Method for screening for  
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity  
; TITLE OF INVENTION: and nucleic acid encoding for UGT  
; FILE REFERENCE: 2139-9"US"  
; CURRENT APPLICATION NUMBER: US/09/376,330  
; CURRENT FILING DATE: 1999-08-18  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 15  
; LENGTH: 1548  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: D. melanogaster UGTT  
US-09-376-330-15

Query Match 6.9%; Score 41; DB 4; Length 1548;  
Best Local Similarity 28.2%; Pred. No. 5.1;  
Matches 11; Conservative 2; Mismatches 26; Indels 0; Gaps 0;

QY 96 PXXXXXXXHXHXXLPDXHXXLXXXXXXYL 134

Db 935 PRQTKRKLPTDLKTDHVSVKLPKQKLPDVAANL 973

##### RESULT 2

US-08-685-466C-4  
; Sequence 4, Application US/08685466C  
; Patent No. 6015705  
; GENERAL INFORMATION:  
; APPLICANT: Jan Willem Vrijbloed and Lubbert Dijkhuizen  
; TITLE OF INVENTION: Method of Modifying the DNA of an  
; TITLE OF INVENTION: Organism, DNA Sequence, Method of  
; TITLE OF INVENTION: Isolation of a Gene Capable of Modifying the DNA of an  
; TITLE OF INVENTION: Organism and Organisms Comprising Modified DNA  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Daniel J. Hulseberg  
; STREET: P.O. Box 2828  
; CITY: Chicago  
; STATE: IL



**THIS PAGE BLANK (USPTO)**

; PRIOR APPLICATION NUMBER: 60/081195  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081838  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082568  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082569  
; PRIOR FILING DATE: 1998-04-21  
; PRIOR APPLICATION NUMBER: 60/082704  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/082797  
; PRIOR FILING DATE: 1998-04-22  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/083495  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083496  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083499  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/083559  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/084366  
; PRIOR FILING DATE: 1998-05-05  
; PRIOR APPLICATION NUMBER: 60/084414  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 60/084639  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084640  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084643  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/085573  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085579  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085580  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085582  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/085700  
; PRIOR FILING DATE: 1998-05-15  
; PRIOR APPLICATION NUMBER: 60/086023  
; PRIOR FILING DATE: 1998-05-18  
; PRIOR APPLICATION NUMBER: 60/086392  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/086486  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/087098  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087208  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087609  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087759  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/087827  
; PRIOR FILING DATE: 1998-06-03  
; PRIOR APPLICATION NUMBER: 60/088025  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088028  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088029  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088033  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088167  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088202  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088212  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088217

; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/088326  
; PRIOR FILING DATE: 1998-06-04  
; PRIOR APPLICATION NUMBER: 60/088655  
; PRIOR FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: 60/088722  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088738  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088740  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088811  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088824  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088825  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088863  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089090  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089908

Query Match 5.3%; Score 31; DB 12; Length 1894;

Best Local Similarity 19.0%; Pred. No. 3.2e+02;

Matches 4; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 155 CXXXXXXXCCCCXXC 175

Db 557 CTTAACAAATAGCATACATC 577

Search completed: November 8, 2002, 19:29:44

Job time : 12.0654 secs

```
;
;
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 559 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 496120
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-836-561-6

Query Match          5.3%: Score 31; DB 10; Length 559;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPD 120
||:|:|:|:|
Db 59 LRVDHSVTLSD 70

RESULT 14
US-09-833-790-429
; Sequence 429, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833.790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-790-429

Query Match          5.3%: Score 31; DB 10; Length 629;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPD 120
||:|:|:|:|
Db 385 LRSESVSRLPD 396

RESULT 15
US-10-052-586-97
; Sequence 97, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```

; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12406  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-12406

Query Match 5.3%; Score 31; DB 10; Length 513;  
Best Local Similarity 29.4%; Pred. No. 84;  
Matches 5; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 104 KXXXXLRTHEHXXVXXLPD 120  
| : : | | |  
Db 338 KTAQTIKSHHNVGGLPE 354

RESULT 11  
US-09-815-242-10987  
; Sequence 10987, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10987  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-815-242-10987

Query Match 5.3%; Score 31; DB 10; Length 523;  
Best Local Similarity 41.7%; Pred. No. 86;  
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 109 LRTEHXXVXXLPD 120  
: : : | | |  
Db 353 IKSHHNVGGLPD 364  
  
RESULT 12  
US-09-854-549-7  
; Sequence 7, Application US/09854549  
; Patent No. US20010034334A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni, et al.  
; TITLE OF INVENTION: Human Extracellular Matrix-1  
; FILE REFERENCE: PF223C1  
; CURRENT APPLICATION NUMBER: US/09/854,549  
; CURRENT FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 09/007,105  
; PRIOR FILING DATE: 1998-01-14  
; PRIOR APPLICATION NUMBER: US 60/035,711  
; PRIOR FILING DATE: 1997-01-16  
; PRIOR APPLICATION NUMBER: US 60/050,113  
; PRIOR FILING DATE: 1997-06-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-854-549-7

Query Match 5.3%; Score 31; DB 10; Length 559;  
Best Local Similarity 50.0%; Pred. No. 92;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 109 LRTEHXXVXXLPD 120  
| : : | | |  
Db 59 LRVDHVSLSLHD 70

RESULT 13  
US-09-836-561-6  
; Sequence 6, Application US/09836561  
; Patent No. US20020038006A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; Corley, Neil C.  
; Guegler, Karl J.  
; TITLE OF INVENTION: HUMAN EXTRACELLULAR MATRIX PROTEINS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/836,561  
; FILING DATE: 16-Apr-2001  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/212,168  
; FILING DATE: <unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0333 US  
; TELECOMMUNICATION INFORMATION:

APPLICANT: Schaffner, Gotthold  
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus  
; FILE REFERENCE: 0652.1800001  
; CURRENT APPLICATION NUMBER: US/09/970,711  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/171,461  
; PRIOR FILING DATE: 1999-01-12  
; PRIOR APPLICATION NUMBER: PCT/EP97/01944  
; PRIOR FILING DATE: 1997-04-18  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 206  
; TYPE: PRT  
; ORGANISM: CELO Virus  
; FEATURE:  
; OTHER INFORMATION: Position: 21134..21754 /gene: L3 /product: L3  
; OTHER INFORMATION: protease  
US-09-970-711-13

Query Match 5.3%; Score 31; DB 10; Length 206;  
Best Local Similarity 42.9%; Pred. No. 33;  
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
||| : |||  
DB 192 LRRETALNALPENH 205

RESULT 8  
US-09-815-242-5490  
; Sequence 5490, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5490  
; LENGTH: 506  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus

Query Match 5.3%; Score 31; DB 10; Length 506;  
Best Local Similarity 29.4%; Pred. No. 83;  
Matches 5; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 104 KXXXLRTEHXVXXLPD 120  
| : : | | |  
DB 331 KTAQTIKSHHNVGGLPE 347

RESULT 9  
US-09-815-242-11555  
; Sequence 11555, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11555  
; LENGTH: 508  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori

Query Match 5.3%; Score 31; DB 10; Length 508;  
Best Local Similarity 41.7%; Pred. No. 83;  
Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPD 120  
||| : |||  
DB 338 IKTHHNVGGLPE 349

RESULT 10  
US-09-815-242-12406  
; Sequence 12406, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 42872  
;; LENGTH: 30  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL031651.32  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5  
;; OTHER INFORMATION: EST\_HUMAN HIT: BF529379.1, EVALUATE 3.00e-03  
US-09-864-761-42872

Query Match 5.3%; Score 31; DB 10; Length 30;  
Best Local Similarity 42.9%; Pred. No. 4.6;  
Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
||:|:|:|:|:|  
Db 17 LRPKHLAQLPPAH 30

RESULT 6  
US-09-864-761-34686  
;; Sequence 34686, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecmica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 34686  
;; LENGTH: 40  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AP000459.2  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 73  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 28  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 83  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 19  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 59  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.3  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 24  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 56  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 30  
;; OTHER INFORMATION: EST\_HUMAN HIT: BE300193.1, EVALUATE 4.00e-08  
;; OTHER INFORMATION: SWISSPROT HIT: Q49412, EVALUATE 4.70e+00  
US-09-864-761-34686

Query Match 5.3%; Score 31; DB 10; Length 40;  
Best Local Similarity 35.7%; Pred. No. 6.2;  
Matches 5; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
:|:|:|:|:|  
Db 1 IHTNHFIILLKDNH 14

RESULT 7  
US-09-970-711-13  
;; Sequence 13, Application US/09970711  
;; Patent No. US20020081279A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Baker, Adam  
;; APPLICANT: Cotten, Matthew  
;; APPLICANT: Chioocca, Susanna  
;; APPLICANT: Kurzbauer, Robert



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:26:04 ; Search time 9.06536 Seconds  
(without alignments)  
464.428 Million cell updates/sec

Title: US-09-840-743-73  
Perfect score: 590  
Sequence: 1 FXXEXPXXPXXXXXXXXXXX.....XXXXXXXXXXLHXXSK 292

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues 92612  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications\_AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	6.3	233	10 US-09-764-855-121	Sequence 121, App
2	37	6.3	3503	9 US-10-108-605-237	Sequence 237, App
3	32	5.4	54	10 US-09-864-761-46893	Sequence 46893, A
4	32	5.4	337	10 US-09-946-143-2	Sequence 2, Appli
5	31	5.3	30	10 US-09-864-761-42872	Sequence 42872, A
6	31	5.3	40	10 US-09-864-761-34686	Sequence 34686, A
7	31	5.3	206	10 US-09-970-711-13	Sequence 13, Appl
8	31	5.3	506	10 US-09-815-242-5490	Sequence 5490, Ap
9	31	5.3	508	10 US-09-815-242-11555	Sequence 11555, A
10	31	5.3	513	10 US-09-815-242-12406	Sequence 12406, A
11	31	5.3	523	10 US-09-815-242-10987	Sequence 10987, A
12	31	5.3	559	10 US-09-854-549-7	Sequence 7, Appli
13	31	5.3	559	10 US-09-836-561-6	Sequence 6, Appli
14	31	5.3	629	10 US-09-833-790-429	Sequence 429, App
15	31	5.3	1894	12 US-10-052-586-97	Sequence 97, Appl
16	30	5.1	141	10 US-09-799-777-16	Sequence 16, Appl
17	30	5.1	145	10 US-09-903-456-41	Sequence 41, Appl
18	30	5.1	148	10 US-09-903-456-51	Sequence 51, Appl
19	30	5.1	199	10 US-09-864-761-33578	Sequence 33578, A

20	30	5.1	199	10 US-09-864-761-33661	Sequence 33661, A
21	30	5.1	209	10 US-09-764-864-800	Sequence 800, App
22	30	5.1	210	10 US-09-811-284-229	Sequence 229, App
23	30	5.1	214	12 US-10-078-929-66	Sequence 66, Appl
24	30	5.1	234	10 US-09-912-020-350	Sequence 350, App
25	30	5.1	273	10 US-09-216-393-12	Sequence 12, Appl
26	30	5.1	283	10 US-09-903-456-85	Sequence 85, Appl
27	30	5.1	292	10 US-09-764-864-1261	Sequence 1261, Ap
28	30	5.1	292	10 US-09-903-456-65	Sequence 65, Appl
29	30	5.1	423	10 US-09-925-300-1233	Sequence 1233, Ap
30	30	5.1	502	10 US-09-323-988D-56	Sequence 56, Appl
31	30	5.1	1422	8 US-08-424-550B-81	Sequence 81, Appl
32	30	5.1	1497	10 US-09-060-854B-2	Sequence 2, Appli
33	30	5.1	4440	12 US-10-052-586-25	Sequence 525, App
34	30	5.1	4679	10 US-09-804-898-2	Sequence 2, Appli
35	29.5	5.0	272	10 US-09-925-118-2	Sequence 2, Appli
36	29.5	5.0	1386	10 US-09-866-582-38	Sequence 38, Appl
37	29	4.9	16	10 US-09-822-698A-51	Sequence 51, Appl
38	29	4.9	44	10 US-09-864-761-37419	Sequence 37419, A
39	29	4.9	60	9 US-10-046-961-17	Sequence 17, Appl
40	29	4.9	70	10 US-09-864-761-49072	Sequence 49072, A
41	29	4.9	80	10 US-09-894-018-97	Sequence 97, Appl
42	29	4.9	98	10 US-09-894-018-101	Sequence 101, App
43	29	4.9	106	10 US-09-894-018-95	Sequence 95, Appl
44	29	4.9	107	10 US-09-894-018-93	Sequence 93, Appl
45	29	4.9	107	10 US-09-894-018-103	Sequence 103, App

ALIGNMENTS

RESULT 1  
US-09-764-855-121  
; Sequence 121, Application US/09764855  
; Patent No. US20020119919A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P1110  
; CURRENT APPLICATION NUMBER: US/09/764,855  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 334  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 121  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (152)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (231)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-855-121

Query Match 6.3%; Score 37; DB 10; Length 233;  
Best Local Similarity 46.2%; Pred. No. 1.8;  
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 110 RTEHVXXLPDXH 122  
| | | | |  
Db 42 ROEHFIKGLPEYH 54

RESULT 2  
US-10-108-605-237  
; Sequence 237, Application US/10108605  
; Patent No. US20020160934A1  
; GENERAL INFORMATION:  
; APPLICANT: Broadus, Julie  
; APPLICANT: Stam, Lynn



**THIS PAGE BLANK (USPTO)**

Best Local Similarity 26.7%; Pred. No. 26;  
Matches 8; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

QY 111 TEHXVXXLPDXHXLXXDXXYYLXIWP 140  
| : : | | | : |  
Db 260 TQGMVNVNDSLVQLTPSDNPYYCLVLPV 289

RESULT 15  
F81423  
L-lactate permease Cj0076c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: F81423  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: F81423  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-560 <PAR>  
A:Cross-references: GB:AL139074; GB:AL111168; NID:q6967505; PIDN:CAB72562.1; PID:q696757  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: lctP; Cj0076c  
C:Superfamily: L-lactate permease

Query Match 5.9%; Score 35; DB 2; Length 560;  
Best Local Similarity 25.0%; Pred. No. 35;  
Matches 8; Conservative 2; Mismatches 22; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXHXLXXDXXYYLXIWP 140  
| : : | | | : |  
Db 243 LSSNHLGAELPDIIISAVVSLAVTTVFLKEWKP 274

Search completed: November 8, 2002, 19:26:48  
Job time : 24.085 secs

A:Residues: 1-1057 <STO>  
A:Cross-references: GB:AE004723; GB:AE004091; NID:g9949067; PIDN:AA06364.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: rne; PA2976  
C:Superfamily: ribonuclease E

Query Match 6.1%; Score 36; DB 2; Length 1057;  
Best Local Similarity 50.0%; Pred. No. 36;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
| : | : | | |  
Db 461 LRTRARIFILPDH 474

RESULT 12

A49335  
1-phosphatidylinositol 4-kinase (EC 2.7.1.67) - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: protein N0795; protein YNL267w  
C:Species: *Saccharomyces cerevisiae*  
C>Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 29-Oct-1999  
C:Accession: A49335; S44456; S60915; S63240; S65117; S39245  
R:Flanagan, C.A.; Schnieders, E.A.; Emerick, A.W.; Kunisawa, R.; Admon, A.; Thorner, J.  
Science 262, 1444-1448, 1993  
A:Title: Phosphatidylinositol 4-kinase: gene structure and requirement for yeast cell viability  
A:Reference number: A49335; MUID:94069320; PMID:8248783  
A:Accession: A49335  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1066 <FA>  
A:Cross-references: GB:L20220; NID:g432495; PIDN:AAA34873.1; PID:g432496  
A:Accession: B49335  
A:Molecule type: protein  
A:Residues: 26-43;152-169;255-267;292-305;332-335;434-452;515-530;543-561;603-609;698-711  
R:Garcia-Bustos, J.F.; Marini, F.; Stevenson, I.; Frei, C.; Hall, M.N.  
EMBO J. 13, 2352-2361, 1994  
A:Title: PIK1, an essential phosphatidylinositol 4-kinase associated with the yeast nucleus  
A:Reference number: S44456; MUID:94252322; PMID:8194527  
A:Accession: S44456  
A:Molecule type: DNA  
A:Residues: 1-1066 <GAR>  
A:Cross-references: EMBL:X76058; NID:g433916; PIDN:CAA53658.1; PID:g433917  
R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.N.; Hegemann, J.H.  
submitted to the EMBL Data Library, October 1995  
A:Description: The sequence of a 24152 bp segment from the left arm of chromosome XIV from *Saccharomyces cerevisiae*  
A:Reference number: S60909  
A:Accession: S60915  
A:Molecule type: DNA  
A:Residues: 1-1066 <SEN>  
A:Cross-references: EMBL:X92494; NID:g1045236; PIDN:CAA63231.1; PID:g1045243  
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63235  
A:Accession: S63240  
A:Molecule type: DNA  
A:Residues: 1-1066 <SEW>  
A:Cross-references: EMBL:Z71543; NID:g1302325; PIDN:CAA96174.1; PID:e239719; PID:g130232  
A:Experimental source: strain S288C  
R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.  
Yeast 12, 505-514, 1996  
A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome XIV from *Saccharomyces cerevisiae*  
A:Reference number: S65111; MUID:96310631; PMID:8740425  
A:Accession: S65117  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 721-960 <SEF>  
A:Cross-references: EMBL:X92494  
C:Genetics:  
A:Gene: SGD; PIK1; PIK41; PIK120  
A:Cross-references: SGD:S0005211; MIPS:YNL267w  
A:Map position: 14L  
C:Keywords: nucleus; phosphotransferase; transmembrane protein

F:156-172/Domain: transmembrane #status predicted <TMM>  
Query Match 6.1%; Score 36; DB 2; Length 1066;  
Best Local Similarity 35.7%; Pred. No. 37;  
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122  
| : | : | | |  
Db 358 LNSDHTSSMPDLH 371

RESULT 13

D69786  
glycoprotein endopeptidase homolog yd1c - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: D69786  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
isch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: D69786  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-229 <KUN>  
A:Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12411.1; PID:g26329  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yd1c  
C:Superfamily: hypothetical protein HI0388

Query Match 5.9%; Score 35; DB 2; Length 229;  
Best Local Similarity 33.3%; Pred. No. 16;  
Matches 7; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 129  
| : | : | | |  
Db 31 LKKNHVRAMPVHSLNDCD 51

RESULT 14

AB0013  
probable exported protein YPO0099 [imported] - *Yersinia pestis* (strain CO92)  
C:Species: *Yersinia pestis*  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AB0013  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AB0013  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-411 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC88964.1; PID:gi5978208; GSPDB:GN00175  
C:Genetics:  
A:Gene: YPO0099

Query Match 5.9%; Score 35; DB 2; Length 411;

Query Match 6.4%; Score 38; DB 2; Length 1540;  
 Best Local Similarity 36.4%; Pred. No. 18;  
 Matches 8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 112 EHXVXXLPDXHXXLXXDXXX 133  
 DB 990 DHDVSTLHDLHNRGVDNVLY 1011

RESULT 7  
 T43274  
 dynein heavy chain, cytosolic - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T43274  
 R:Yamamoto, A.; West, R.R.; McIntosh, J.R.; Hiraoka, Y.  
 J. Cell Biol. 145, 1233-1250, 1999  
 A:Title: A cytoplasmic dynein heavy chain is required for oscillatory nuclear movement  
 A:Reference number: 222383; MUID:9296615; PMID:10366596  
 A:Accession: T43274  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4196 <YAM>  
 A:Cross-references: EMBL:AB006784; NID:g2351080; PIDN:BAA22056.1; PTD:g2351081  
 A:Experimental source: strain CRL152  
 C:Genetics:  
 A:Gene: dhcl  
 A:Map position: 1  
 C:Superfamily: dynein heavy chain, cytosolic

Query Match 6.4%; Score 38; DB 2; Length 4196;  
 Best Local Similarity 36.8%; Pred. No. 45;  
 Matches 7; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 111 TEHXVXXLPDXHXXLXXD 129  
 DB 1140 SEHLLNLDTHSSLTVD 1158

RESULT 8  
 T26148  
 hypothetical protein W04D2.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T26148  
 R:Lennard, N.  
 submitted to the EMBL Data Library, June 1996  
 A:Reference number: Z20161  
 A:Accession: T26148  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-624 <WIL>  
 A:Cross-references: EMBL:Z75552; PIDN:CAA99942.1; GSPDB:GN00023; CESP:W04D2.4  
 A:Experimental source: clone W04D2  
 C:Genetics:  
 A:Gene: CESP:W04D2.4  
 A:Map position: 5  
 A:Introns: 23/3; 112/3; 183/2; 402/2; 495/1; 523/3

Query Match 6.3%; Score 37; DB 2; Length 624;  
 Best Local Similarity 64.3%; Pred. No. 14;  
 Matches 9; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 109 LRTEHXXVXXLPDXH 122  
 DB 90 LRKIEHDV--LPDVH 101

RESULT 9  
 B69305  
 succinoglycan biosynthesis regulator (exsB) homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
 C:Accession: B69305  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: B69305  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-239 <KLE>  
 A:Cross-references: GB:AF001074; GB:AE000782; NID:g2689397; PIDN:AAB90792.1; PTD:g265  
 C:Superfamily: conserved hypothetical protein H1191

Query Match 6.1%; Score 36; DB 2; Length 239;  
 Best Local Similarity 50.0%; Pred. No. 9.7;  
 Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 129 DXXXVLLXIWP 140  
 DB 141 DPAVLANIWP 152

RESULT 10  
 AH0853  
 DNA mismatch repair protein [imported] - Salmonella enterica subsp. enterica serovar  
 C:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: This species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AH0853  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
 , S.; Moulé, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, S  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AH0853  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-855 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD06015.1; PID:g16503984; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STX3033  
 C:Superfamily: DNA mismatch repair protein mutS

Query Match 6.1%; Score 36; DB 2; Length 855;  
 Best Local Similarity 35.0%; Pred. No. 30;  
 Matches 7; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 110 RTEHXXVXXLPDXHXXLXXD 129  
 DB 369 RMRHAFQQLPELHAQLETVD 388

RESULT 11  
 H83273  
 ribonuclease E PA2976 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H83273  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; L  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: H83273  
 A:Status: preliminary  
 A:Molecule type: DNA



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	146	24.7	234	2	T48454	hypothetical prote	
2	91	15.4	1207	2	D84781	hypothetical prote	
3	77	13.1	917	2	T05430	hypothetical prote	
4	41	6.9	1548	2	S54723	UDP-glucose-glycop	
5	38	6.4	1446	2	S73013	polyketide synthas	
6	38	6.4	1540	2	H87203	polyketide synthas	
7	38	6.4	4196	2	T43274	dynein heavy chain	
8	37	6.3	624	2	T26148	hypothetical prote	
9	36	6.1	239	2	B69305	succinoglycan bios	
10	36	6.1	855	2	A80853	DNA mismatch repai	
11	36	6.1	1057	2	H83273	ribonuclease E PA2	
12	36	6.1	1066	2	A49335	1-phosphatidylinos	
13	35	5.9	229	2	D65786	glycoprotein endop	
14	35	5.9	411	2	AB0013	probable exported	
15	35	5.9	560	2	F81423	L-lactate permease	
16	35	5.9	762	2	H90451	hypothetical prote	
17	34	5.8	249	2	T10358	protein cg30 - Org	
18	34	5.8	250	2	B72454	probable thioredox	
19	34	5.8	298	2	D87450	transcription regu	
20	34	5.8	322	2	H81213	hypothetical prote	
21	34	5.8	322	2	B81790	hypothetical regu	
22	34	5.8	335	2	A87451	transcription regu	
23	34	5.8	459	2	D71331	probable UDP-N-ace	
24	34	5.8	499	2	F69833	monooxygenase homo	
25	34	5.8	510	2	E70321	GMP synthase (glut	
26	34	5.8	560	2	E55083	Putative L-lactate	
27	34	5.8	562	2	F83050	L-lactate permease	
28	34	5.8	567	2	C75340	probable L-lactate	
29	34	5.8	586	2	F93540	formate dehydrogen	
30	34	5.8	787	2	A13556	formate dehydrogen	

**THIS PAGE BLANK (USPTO)**

```

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.F.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: CONSISTS OF A SIGMA FACTOR AND THE RNAP CORE ENZYME WHICH
CC IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, 1 BETA' CHAIN AND 1
CC OMEGA CHAIN (BY SIMILARITY).
CC -!- DOMAIN: THE AMINO-TERMINAL PORTION IS INVOLVED IN THE ASSEMBLY OF
CC CORE RNAP. WHEREAS THE C-TERMINAL IS INVOLVED IN INTERACTION WITH
CC TRANSCRIPTIONAL REGULATORS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003951; AAF83986.1; -
DR HSSP; P00574; ICGO.
DR InterPro; IPR001700; RNA_pol_A_bac.
DR Pfam; PF01000; RNA_pol_A_bac; 1.
DR Pfam; PF03118; RNA_pol_A_CTD; 1.
DR ProDom; PD001179; RNA_pol_A_bac; 1.
KW Transferase; Transcription; DNA-directed RNA polymerase;
KW -Complete proteome.
SQ SEQUENCE 332 AA; 36554 MW; 98BDE535F753A577 CRC64;

Query Match 5.6%; Score 33; DB 1; Length 332;
Best Local Similarity 35.7%; Pred. No. 17;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 109 LRTEHXXVXXLPDXH 122
Db 115 IKTDHNVETINDGH 128
:::| | : | |

```

Search completed: November 8, 2002, 19:27:16  
Job time : 15.4967 secs



```
DR TIGRFAMS; TIGR01080; rplX_A_E; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; 1.
KW Ribosomal protein.
SQ SEQUENCE 142 AA; 16075 MW; B496362F0BFEC354 CRC64;

Query Match
Best Local Similarity 5.6%; Score 33; DB 1; Length 142;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 109 LRTEHVVXXLPD 119
      ||||| : :|
Db 38 LRTKHGIRAIP 48

RESULT 13
TDXH_PVRAB STANDARD; PRT; 216 AA.
AC Q9U2V4:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable peroxiredoxin.
GN PAB1673.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN# [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GE5 / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. TDXH SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AP000005; BAA30317.1; -.
CC HSSP; P30041; 1PRX.
CC InterPro; IPR000866; AhpC-TSA.
CC Pfam; PF00578; AhpC-TSA; 1.
CC Antioxidant; Complete proteome.
FT ACT_SITE 46 46 BY SIMILARITY.
SQ SEQUENCE 216 AA; 24758 MW; A95DF14BF3874CA9 CRC64;

Query Match
Best Local Similarity 5.6%; Score 33; DB 1; Length 216;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 109 LRTEHVVXXLPD 120
      ::||| |||
Db 13 VKTTHGVKLPD 24

RESULT 15
RPOA_XYLFA STANDARD; PRT; 332 AA.
AC Q9PE52:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase alpha chain (EC 2.7.7.6) (Transcriptase
DE alpha chain) (RNA polymerase alpha subunit).
GN RPOA OR XF1176.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
```

```
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 314 334 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 440 460 POTENTIAL.
FT TRANSMEM 538 558 POTENTIAL.
SQ SEQUENCE 560 AA; 58920 MW; 8ECB9F626FFC40A CRC64;

Query Match 5.8%; Score 34; DB 1; Length 560;
Best Local Similarity 30.4%; Pred. No. 17; Mismatches 15; Indels 0; Gaps 0;
Matches 7; Conservative 1;

Qy 118 LPDXXHXXLXXXIXIXWP 140
||| | | |
Db 249 LPDITSALVSIVSLALFLKVRP 271

RESULT 11
CYAA_NEUCR
ID CYAA_NEUCR STANDARD; PRT; 2300 AA.
AC Q01631;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl
cyclase).
DE CR-1 OR NAC.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92000795; PubMed=1680356;
RA Kore-Eda S., Murayama T., Uno I.;
RT "Isolation and characterization of the adenylate cyclase structural
gene of Neurospora crassa.";
RL Jpn. J. Genet. 66:317-334(1991).
CC -1- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR
METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,
CAMP.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- SIMILARITY: BELONGS TO ADENYL CYCLASE CLASS-3 FAMILY.
CC -1- SIMILARITY: CONTAINS 23 LEUCINE-RICH REPEATS (LRR).
CC -1- SIMILARITY: CONTAINS 1 PP2C-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00909; BAA00755.1; -.
DR InterPro: IPR001054; G.cyclase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_typ.
DR InterPro: IPR001932; PP2C-like.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF00211; guanylate_cyc; 1.
DR Pfam: PF00481; PP2C; 1.
DR Pfam: PF00560; LRR; 14.
DR Pfam: PF00788; RA; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00044; CYC; 1.
DR SMART; SM00370; LRR; 4.
DR SMART; SM00369; LRR_TYP; 2.
DR SMART; SM00332; PP2Cc; 1.
```

```
DR SMART; SM00314; RA; 1.
DR PROSITE; PSS0125; GUANYLATE_CYCLASES_2; 1.
KW Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.
FT REPEAT 867 890 LRR 1.
FT REPEAT 892 914 LRR 2.
FT REPEAT 915 938 LRR 3.
FT REPEAT 939 961 LRR 4.
FT REPEAT 962 986 LRR 5.
FT REPEAT 988 1008 LRR 6.
FT REPEAT 1009 1031 LRR 7.
FT REPEAT 1033 1055 LRR 8.
FT REPEAT 1056 1079 LRR 9.
FT REPEAT 1081 1097 LRR 10.
FT REPEAT 1098 1120 LRR 11.
FT REPEAT 1122 1142 LRR 12.
FT REPEAT 1143 1165 LRR 13.
FT REPEAT 1166 1188 LRR 14.
FT REPEAT 1189 1211 LRR 15.
FT REPEAT 1213 1234 LRR 16.
FT REPEAT 1349 1369 LRR 17.
FT REPEAT 1373 1396 LRR 18.
FT REPEAT 1398 1420 LRR 19.
FT REPEAT 1422 1445 LRR 20.
FT REPEAT 1447 1469 LRR 21.
FT REPEAT 1474 1497 LRR 22.
FT REPEAT 1513 1538 LRR 23.
FT DOMAIN 1564 1829 PP2C-LIKE.
FT DOMAIN 1830 2300 CATALYTIC.
FT DOMAIN 49 52 POLY-SER.
FT DOMAIN 204 221 POLY-GLN.
SQ SEQUENCE 2300 AA; 254752 MW; 52E79B90B6B17A7B CRC64;

Query Match 5.8%; Score 34; DB 1; Length 2300;
Best Local Similarity 31.6%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 104 KXXXXLRTXHXVXXLPDXH 122
| | | | |
Db 462 KTSSSARSGHSHVHPGHH 480

RESULT 12
RL26_CAEEL
ID RL26_CAEEL STANDARD; PRT; 142 AA.
AC Q19869;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L26.
GN RPL-26 OR F28C6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Burton J.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z68315; CAA92674.1; -.
DR WormPep; F28C6.7; CE03278.
DR InterPro: IPR000302; KOW_motif.
DR Pfam; PF00467; KOW; 1.
```

OY 111 TEHXVXLDPD 120  
||||| :||  
Db 224 TEHSVGFVPD 233

```
RESULT 9
GUA_AQUAE
ID GUAA_AQUAE STANDARD: PRT; 510 AA.
AC O66601;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine
  amidotransferase) (GMP synthetase).
GN GUAA OR AQ.236.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
  Graham D.E., Overbeek R., Sneed M.A., Keller M., Aufay M., Huber R.,
  Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
  aeolicus.";
RL Nature 392:353-358(1998).
CC -!- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
  H(2)O = AMP + diphosphate + GMP + L-glutamate.
CC -!- PATHWAY: GMP biosynthesis.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE
  FAMILY.
CC -!- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000679; AAC06558.1; -.
DR HSSP; P04079; 1GPM.
DR InterPro; IPR002385; Anth_synthII.
DR InterPro; IPR000391; GATase_1.
DR InterPro; IPR001674; GMP_synt_C.
DR InterPro; IPR004739; GMPsynthase_N.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00958; GMP_synt_C; 1.
DR PRINTS; PR00097; ANTSYNTHASEII.
DR PRINTS; PR00096; GATASE.
DR TIGRfams; TIGR00884; guaA_Cterm; 1.
DR TIGRfams; TIGR00888; guaA_Nterm; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
DR Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
KW Glutamine amidotransferase; Complete proteome.
FT DOMAIN 1 189 GLUTAMINE AMIDOTRANSFERASE.
FT DOMAIN 227 386 GMP-BINDING (BY SIMILARITY).
FT ACT_SITE 82 82 GATASE (BY SIMILARITY).
FT ACT_SITE 169 169 GATASE (BY SIMILARITY).
FT ACT_SITE 171 171 GATASE (BY SIMILARITY).
FT NP_BIND 223 229 ATP (BY SIMILARITY).
SQ SEQUENCE 510 AA; 57858 MW; 2855029D58D1A667 CRC64;

Query Match 5.8%; Score 34; DB 1; Length 510;
Best Local Similarity 35.3%; Pred. NO. 16;
Matches 6; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 104 KXXXLRTHEHXVXLDPD 120
```

```
Db 335 KGSAKIKTHNVGGLPE 351
| :| | | | |
| :| | | | |

RESULT 10
GLCA_ECOLI
ID GLCA_ECOLI STANDARD: PRT; 560 AA.
AC Q46839;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycolate permease glca.
GN GLCA OR B2975.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
  Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP CHARACTERIZATION
RX STRAIN=K12 / MC4100;
RX MEDLINE=21178909; PubMed=11283302;
RA Nunez M.F., Pellicer M.T., Badia J., Aguilar J., Baldoma L.;
RT "The gene yghK linked to the glc operon of Escherichia coli encodes a
  permease for glycolate that is structurally and functionally similar
  to L-lactate permease.";
RT Microbiology 147:1069-1077(2001).
RN [3]
RP CHARACTERIZATION
RX STRAIN=K12 / MC4100;
RX MEDLINE=21645912; PubMed=11785976;
RA Nunez M.F., Kwon O., Wilson T.H., Aguilar J., Baldoma L., Lin E.C.C.;
RT "Transport of L-lactate, D-lactate, and glycolate by the LldP and GlcA
  membrane carriers of Escherichia coli.";
RL Biochem. Biophys. Res. Commun. 290:824-829(2002).
CC -!- FUNCTION: Transports glycolate across the membrane. Can also
  transport L-lactate and D-lactate. Seems to be driven by a proton
  motive force.
CC -!- ENZYME REGULATION: Inhibited by carbonyl cyanide m-
  chlorophenylhydrazine (CCCP).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
  (Probable).
CC -!- INDUCTION: By glycolate.
CC -!- SIMILARITY: BELONGS TO THE LACTATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U28377; AAA69142.1; -.
DR EMBL; AE000380; AAC76011.1; -.
DR EcoGene; EG12995; glca.
DR InterPro; IPR003804; Lactate_perm.
DR Pfam; PF02652; Lactate_perm; 1.
DR TIGRfams; TIGR00795; lctP; 1.
DR Transmem 14 34 POTENTIAL.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 131 151 POTENTIAL.
```

```
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; AAC59088.1; -.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Zinc-finger.
FT ZN_FING 8 66 RING-TYPE.
SQ SEQUENCE 249 AA; 28295 MW; 03B17D50BA9EE4B CRC64;

Query Match 5.8%; Score 34; DB 1; Length 249;
Best Local Similarity 50.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
DB 193 LRAERAVKALADAH 206

RESULT 7
TXDH_AERPE
ID TXDH_AERPE STANDARD; PRT; 250 AA.
AC QY9YLO;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable peroxiredoxin.
GN APE2278.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. TDXH SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000064; BAA81290.1; -.
DR HSP; P30041; IPRX.
DR InterPro; IPR000866; AhpC-TSA.
DR Pfam; PF00578; AhpC-TSA; 1.
KW Antioxidant; Complete proteome.
FT ACT_SITE 50 50 BY SIMILARITY.
SQ SEQUENCE 250 AA; 28703 MW; 0457F2852D051E7A CRC64;
```

```
Query Match 5.8%; Score 34; DB 1; Length 250;
Best Local Similarity 50.0%; Pred. No. 7.5;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 111 TEHXVXXLPDXH 122
DB 19 TDHGVIKLPDHY 30

RESULT 8
MURF_TREPA
ID MURF_TREPA STANDARD; PRT; 459 AA.
AC O83401;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylmuramoylalanine-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAC-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
GN MURF OR TP0386.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete";
RL Science 281:375-388(1998).
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine - ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001217; AAC65370.1; -.
DR HSP; P11880; IGG4.
DR TIGR; TP0386; -.
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR Pfam; PF01225; Mur_ligase; 1.
DR Pfam; PF02875; Mur_ligase_C; 1.
DR TIGRfams; TIGR01143; murF; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW ATP-binding; Complete proteome.
FT NP_BIND 121 127 ATP (POTENTIAL).
SQ SEQUENCE 459 AA; 50234 MW; 23EA5571ED95F332 CRC64;

Query Match 5.8%; Score 34; DB 1; Length 459;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
CC      step. This protein has a weak ATPase activity.
CC      -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC      -!- CAUTION: Ref.1 sequence differs from that shown due to
CC      frameshifts in positions 353, 379, 394, 422 and 425.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U16303; AA80578.1; ALT_FRAME.
DR      EMBL; M18965; AA27167.1; ALT_FRAME.
DR      EMBL; AE008832; AAL21789.1; -.
DR      EMBL; AL627276; CAD06015.1; -.
DR      PIR; A28668; A28668.
DR      StyGene; SG10237; mutS.
DR      InterPro; IPR000432; Muts_C.
DR      InterPro; IPR002863; Muts_N.
DR      Pfam; PF00488; Muts_C; 1.
DR      Pfam; PF01624; Muts_N; 1.
DR      SMART; SM00534; MutsSac; 1.
DR      TIGRFAMS; TIGR01070; mutS1; 1.
DR      PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW      DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT      NP_BIND 616 623      ATP (POTENTIAL).
FT      MUTAGEN 622 622      K->A: DEFECTIVE IN MISMATCH REPAIR;
FT      ATPASE ACTIVITY? REDUCED 6 FOLD.
FT      TA -> S (IN REF. 1).
FT      CONFLICT 361 362
SQ      SEQUENCE 855 AA; 95406 MW; 804BEB13BCD2F6F CRC64;

Query Match          6.1%; Score 36; DB 1; Length 855;
Best Local Similarity 35.0%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 110 RTEHXVXXLPDXHXXLXXXD 129
      | | | | | | | |
DB 369 RMRHAFQQLPELHAQLETVD 388

RESULT 5
PIKL_YEAST
ID      PIKL1_YEAST          STANDARD;          PRT; 1066 AA.
AC      P39104;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-OCT-1996 (Rel. 34, Last annotation update)
DE      Phosphatidylinositol 4-kinase PIK1 (BC 2.7.1.67) (PI4-kinase)
DE      (PtdIns-4-kinase).
GN      PIK1 OR YNL267W OR N0795.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=JK9-3D;
RX      MEDLINE=94252322; PubMed=8194527;
RA      Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N.;
RT      "PIK1, an essential phosphatidylinositol 4-kinase associated with the
RT      yeast nucleus.";
RL      EMBO J. 13:2352-2361(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94069320; PubMed=8248783;
RA      Flanagan C.A., Schmieders E.A., Emerick A.W., Kunisawa R., Admon A.,
RT      Thorne J.;
RT      "Phosphatidylinositol 4-kinase: gene structure and requirement for
RT      yeast cell viability.";
RL      Science 262:1444-1448(1993).
RN      [3]
```

```
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288C / FY1679;
RX      MEDLINE=96310631; PubMed=8740425;
RA      Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;
RT      "The sequence of a 24,152 bp segment from the left arm of chromosome
RT      XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
RT      genes.";
RL      Yeast 12:505-514(1996).
CC      -!- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST
CC      COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER
CC      INOSITOL-1,4,5 -TRISPHOSPHATE. PIK1 IS PART OF A NUCLEAR
CC      PHOSPHOINOSITIDE CYCLE AND COULD CONTROL CYTOKINESIS THROUGH
CC      THE ACTIN CYTOSKELETON.
CC      -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-ID-myo-inositol = ADP +
CC      1-phosphatidyl-ID-myo-inositol 4-phosphate.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X76058; CAA53658.1; -.
DR      EMBL; L20220; AAA34873.1; -.
DR      EMBL; X92494; CAA63231.1; -.
DR      EMBL; Z71543; CAA96174.1; -.
DR      PIR; A49335; A49335.
DR      PIR; S44456; S44456.
DR      SGD; S0005211; PIK1.
DR      InterPro; IPR000403; PI3_PI4_Kinase.
DR      Pfam; PF00454; PI3_PI4_kinase; 1.
DR      SMART; SM00146; PI3Kc; 1.
DR      PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR      PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR      PROSITE; PS02090; PI3_4_KINASE_3; 1.
KW      Transferase; Kinase; Nuclear protein.
FT      DOMAIN 793 1041      PI3K/PI4K.
SQ      SEQUENCE 1066 AA; 119922 MW; 7666979CA14B1CB5 CRC64;

Query Match          6.1%; Score 36; DB 1; Length 1066;
Best Local Similarity 35.7%; Pred. No. 12;
Matches 5; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 109 LRTEHXVXXLPDXH 122
      | : | | | |
DB 358 LNSDHTSSMPDLH 371

RESULT 6
VCG3_NPVOP
ID      VCG3_NPVOP          STANDARD;          PRT; 249 AA.
AC      O10339;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Zinc-finger protein CG30.
GN      CG30.
OS      Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMPNV).
OC      Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC      Nucleopolyhedrovirus.
OX      NCBI_TaxID=164623;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97271300; PubMed=9126251;
RA      Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RT      Rohrmann G.F.;
RT      "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT      polyhedrosis virus genome.";
RL      Virology 229:381-399(1997).
```

CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1  
CC DOMAINS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.  
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ250725; CAC20419.1; -.  
DR Genew; HGNC:14605; ADAMTS12.  
DR MIM; 606184; -.  
DR InterPro; IPR001762; Disintegrin.  
DR InterPro; IPR002870; Pep\_M12B\_propep.  
DR InterPro; IPR001590; Repolysin.  
DR InterPro; IPR000884; TSPL.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF00090; tsp\_1; 6.  
DR Pfam; PF01421; Repolysin; 1.  
DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
DR SMART; SMO0209; TSPL; 8.  
DR PROSITE; PS02015; ADAM\_MEPRO; 1.  
DR PROSITE; PS04027; DISINTEGRIN\_1; FALSE\_NEG.  
DR PROSITE; PS00092; TSPL; 2.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
KW Repeat; Extracellular matrix.  
FT SIGNAL 1 25 POTENTIAL.  
FT PROPEP 26 240 BY SIMILARITY.  
FT CHAIN 241 1593 ADAMTS-12.  
FT DOMAIN 465 544 DISINTEGRIN-LIKE.  
FT DOMAIN 545 596 TSP TYPE-1 1.  
FT DOMAIN 597 700 CYS-RICH.  
FT DOMAIN 701 826 SPACER 1.  
FT DOMAIN 827 881 TSP TYPE-1 2.  
FT DOMAIN 886 943 TSP TYPE-1 3.  
FT DOMAIN 947 995 TSP TYPE-1 4.  
FT DOMAIN 996 1315 SPACER 2.  
FT DOMAIN 1316 1364 TSP TYPE-1 5.  
FT DOMAIN 1367 1423 TSP TYPE-1 6.  
FT DOMAIN 1426 1471 TSP TYPE-1 7.  
FT DOMAIN 1426 1471 TSP TYPE-1 8.  
FT DOMAIN 302 305 POLY-GLU.  
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).  
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 393 393 BY SIMILARITY.  
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).  
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1593 AA; 177545 MW; 079F48E63BD83A3 CRC64;

Query Match 6.3%; Score 37; DB 1; Length 1593;  
Best Local Similarity 46.2%; Pred. No. 11;  
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 110 RTEHVXXLPDXH 122  
DB 41 ROEHFIKGLPEYH 53  
RESULT 4  
ID MUTS\_SALTY STANDARD; PRT; 855 AA.  
AC P10339;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA mismatch repair protein mutS.  
GN MUTS OR STM2909 OR STY3033.  
OS Salmonella typhimurium, and  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602, 601;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.  
RC SPECIES=S.typhimurium;  
RX MEDLINE=88086868; PubMed=3275609;  
RA Haber L.T., Pang P.P., Sobell D.I., Mankovich J.A., Walker G.C.;  
RT "Nucleotide sequence of the Salmonella typhimurium mutS gene required  
RT for mismatch repair: homology of MutS and HexA of Streptococcus  
RT pneumoniae.";  
RL J. Bacteriol. 170:197-202(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhi; STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
RN [4]  
RP MUTAGENESIS, AND ATP-BINDING.  
RC SPECIES=S.typhimurium;  
RX MEDLINE=91330898; PubMed=1651234;  
RA Haber L.T., Walker G.C.;  
RT "Altering the conserved nucleotide binding motif in the Salmonella  
RT typhimurium MutS mismatch repair protein affects both its ATPase and  
RT mismatch binding activities.";  
RL EMBO J. 10:2707-2715(1991).  
RN [5]  
RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 394-425.  
RC SPECIES=S.typhimurium;  
RX MEDLINE=94087727; PubMed=7903399;  
RA Claverie J.-M.;  
RT "Detecting frame shifts by amino acid sequence comparison.";  
RL J. Mol. Biol. 234:1140-1157(1993).  
CC -!- FUNCTION: This protein is involved in the repair of mismatches in  
CC DNA. It is possible that it carries out the mismatch recognition

```

KW Glycoprotein. 1 22
FT SIGNAL 23 1548
FT CHAIN
FT
FT UDP-GLUCOSE:GLYCOPROTEIN
FT GLUCOSYLTRANSFERASE.
FT CARBOHYD 181 181. N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 864 864 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SITE 1545 1548 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 1548 AA; 174465 MW; 95D6849961622DB6 CRC64;

Query Match 6.9%; Score 41; DB 1; Length 1548;
Best Local Similarity 28.2%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 11; Conservative 2; Mismatches 26;

QY 96 PRXXXXXXXLTHKXVXLPDXHXXLXXDXXXL 134
DB 935 PRQTKTRFKLPDLKTHSVVVKLPKQKLPHEFDVAAYL 973

RESULT 2
ID DYHC_SCHPO STANDARD; PRT; 4196 AA.
AC O13290; Q9P6L0; Q9UTP8;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Dynein heavy chain, cytosolic (DYHC).
GN DHC1 OR SPAC30C2.01C OR SPAC1093.06C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CRL152;
RA Yamamoto A., West R.R., McIntosh J.R., Hiraoka Y.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouris J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymoprez B.,
RA Weidjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer C., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Hunt C., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe.";
RN Nature 415:871-880 (2002).
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES. REQUIRED FOR NUCLEAR MOVEMENT
CC DURING MEIOTIC PROPHASE.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF

```

```

CC INTERMEDIATE AND LIGHT CHAINS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. PROBABLY BINDS INDIRECTLY TO
CC THE INNER PLASMA MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB006784; BAA22056.1; -
CC DR EMBL; AL355652; CAB90788.1; -
CC DR EMBL; AL132839; CAB60251.1; -
CC KW Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
CC FT DOMAIN 1217 1252 COILED COIL (POTENTIAL).
CC FT DOMAIN 1984 2012 MICROTUBULE-BINDING (POTENTIAL).
CC FT DOMAIN 3315 3403 COILED COIL (POTENTIAL).
CC FT DOMAIN 3649 3666 COILED COIL (POTENTIAL).
CC FT NP_BIND 1890 1897 ATP (POTENTIAL).
CC FT NP_BIND 2169 2176 ATP (POTENTIAL).
CC FT NP_BIND 2174 2181 ATP (POTENTIAL).
CC FT NP_BIND 2520 2527 ATP (POTENTIAL).
CC SQ SEQUENCE 4196 AA; 484308 MW; 8F10AE370184FC0C CRC64;

Query Match 6.4%; Score 38; DB 1; Length 4196;
Best Local Similarity 36.8%; Pred. No. 17; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 10;

QY 111 TEHVXVXLPDXHXXLXXD 129
DB 1140 SEHLLNELLDDTHSSLSTVD 1158

RESULT 3
ID AT12_HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RA MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueuel J.M., Fernandez P.L., Lopez-Otin C.;
RA "Identification, characterization, and intracellular processing of
RA ADAM-TS12, a novel human disintegrin with a complex structural
RA organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940 (2001).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2002, 19:24:14 ; Search time 10.4967 Seconds  
(without alignments)  
1153.796 Million cell updates/sec

Title: US-09-840-743-73

Perfect score: 590

Sequence: 1 PXXEXPPXPPXXXXXXX.....XXXXXXXXXXLHXXXXX 292

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	41	6.9	1548	1	UGGG_DROME	Q09332 drosophila
2	38	6.4	1496	1	DYHC_SCHPO	Q13290 schizosacch
3	37	6.3	1593	1	ATL2_HUMAN	P58397 homo sapien
4	36	6.1	855	1	MOTS_SALTY	P10339 salmonella
5	36	6.1	1066	1	PIK1_YEAST	P39104 saccharomyc
6	34	5.8	249	1	VCG3_NPVOP	O10339 orgyia pseu
7	34	5.8	250	1	TDXH_AERPE	Q9y910 aeropyrum p
8	34	5.8	459	1	MURE_TREPA	O83401 treponema p
9	34	5.8	510	1	GUAA_AQUAE	O66601 aquifex aeo
10	34	5.8	560	1	GLCA_ECOLI	Q46839 escherichia
11	34	5.8	2300	1	CYAA_NEUCR	Q01631 neurospora
12	33	5.6	142	1	RL26_CAEEL	Q19869 caenorhabdi
13	33	5.6	216	1	TDXH_PYRAB	Q9uzv4 pyrococcus
14	33	5.6	216	1	TDXH_PYRHO	O58966 pyrococcus
15	33	5.6	332	1	RPOA_XYLFA	Q9pe22 xyliella fas
16	33	5.6	355	1	DVRI_BRARE	P35621 brachydanio
17	33	5.6	380	1	T701_FREDI	Q00461 fremyella d
18	33	5.6	521	1	DRTS_TRYCR	Q27793 trypanosoma
19	33	5.6	1023	1	TSCC_PSEAM	P55019 pseudopleur
20	33	5.6	1098	1	RP16_DROME	Q9qgn0 drosophila
21	33	5.6	1275	1	RFBC_MYXXA	Q50864 myxococcus
22	33	5.6	3924	1	ANK2_HUMAN	Q01484 homo sapien
23	32.5	5.5	293	1	TYRC_ZYMMO	Q04983 zymomonas m
24	32.5	5.5	517	1	SEST_CAEEL	Q9n4d6 caenorhabdi
25	32	5.4	190	1	RHOL_DROME	Q24192 drosophila
26	32	5.4	232	1	RFAY_SALTY	P26472 salmonella
27	32	5.4	430	1	MESJ_HABIN	P44689 haemophilus
28	32	5.4	490	1	YE0J_SCHPO	Q10990 schizosacch
29	32	5.4	492	1	WZXC_ECOLI	P77377 escherichia
30	32	5.4	537	1	CNEI_HUMAN	Q99829 homo sapien
31	32	5.4	555	1	HEXA_CAEEL	Q22492 caenorhabdi
32	32	5.4	578	1	IL0R_HUMAN	Q13651 homo sapien
33	32	5.4	775	1	PFL_STRMU	Q59934 streptococc

#### ALIGNMENTS

##### RESULT 1

UGGG_DROME	STANDARD;	PRT;	1548 AA.
AC Q09332;			
DT 01-NOV-1997 (Rel. 35, Created)			
DT 01-NOV-1997 (Rel. 35, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DE UDP-glucose:glycoprotein glucosyltransferase precursor (EC 2.4.1.-)			
DE (UDP-Glc:glycoprotein glucosyltransferase) (dUGT).			
GN UGT OR UGGG.			
OS Drosophila melanogaster (Fruit fly).			
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;			
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
OX NCBI_TaxID=7227;			
RN [1]			
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-37.			
RC TISSUE=Embryo;			
RX MEDLINE=95246722; PubMed=7729408;			
RA Parker C.G., Fessler L.I., Nelson R.E., Fessler J.H.;			
RT "Drosophila UDP-glucose:glycoprotein glucosyltransferase: sequence			
RT and characterization of an enzyme that distinguishes between			
RT denatured and native proteins."			
RL EMBO J. 14:1294-1303(1995).			
CC -!- FUNCTION: UNFOLDED, DENATURED GLYCOPROTEINS ARE SUBSTANTIALLY			
CC BETTER SUBSTRATES FOR GLYCOSYLATION BY THIS ENZYME THAN ARE THE			
CC CORRESPONDING NATIVE PROTEINS. THIS PROTEIN AND TRANSIENT			
CC GLYCOSYLATION MAY BE INVOLVED IN MONITORING AND/OR ASSISTING THE			
CC FOLDING AND ASSEMBLY OF NEWLY MADE GLYCOPROTEINS, IN ORDER TO			
CC IDENTIFY GLYCOPROTEINS THAT NEED ASSISTANCE IN FOLDING FROM			
CC CHAPERONES.			
CC -!- COFACTOR: REQUIRES CALCIUM AND MANGANESE IONS FOR ACTIVITY.			
CC -!- PATHWAY: GLYCOSYLATION.			
CC -!- SUBUNIT: MONOMER.			
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.			
CC THE EARLIEST EMBRYOS, INCREASING AT 6-8 HRS WITH A MAXIMUM AT 10-			
CC 12 HRS. LEVELS DECREASE THEREAFTER AND ARE NOT DETECTED IN 18-20			
CC HRS EMBRYOS AND FIRST INSTAR LARVAE BUT IS DETECTED AGAIN AT			
CC SECOND INSTAR TO PUPATION.			
CC -!- SIMILARITY: SOME, TO YEAST KRE5, S.TYPHIMURIUM RPAJ AND E.COLI			
CC REAI PROTEINS.			
CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC the European Bioinformatics Institute. There are no restrictions on its			
CC use by non-profit institutions as long as its content is in no way			
CC modified and this statement is not removed, usage by and for commercial			
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC or send an email to license@isb-sib.ch).			
CC -----			
CC EMBL; U20554; AAA85850.1; -			
CC FlyBase; FBgn0014075; Ugt.			
CC InterPro; IPR002495; GT 8.			
CC Pfam; PF01501; Glyco_transf_8; 1.			
CC Signal; Transferase; Glycosyltransferase; Endoplasmic reticulum;			
KW			



**THIS PAGE BLANK (USPTO)**

```

Db 301 TTGAACAGATAGTACAACTACTGGACATGAAATCCAGAGCGAGATCTGACAAAAGTA 360
QY 1480 tgcagagcattgactcgtctgctgttaattgcagcgaagactactgaacaaatgatg 1539
Db 361 TCGAGAGCAATTATGGACTCGTCTGCTGTTAAATGCGACGAAGCTACTGAACAAAATGATG 420
QY 1540 gcagcagacaagtgtcttgaggctcgaccttaacaaactctcagcagaacccctcca 1599
Db 421 GCAGCAGCAAGATGTTCTGGAGTTCGACCTTAACAAAACCTCCTGCGAAGCCCTCCA 480
QY 1600 aaagga 1605
Db 481 NAAGGA 486

RESULT 7
AV545791/c
LOCUS AV545791 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION AV545791 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
CDNA clone RZL01d12F 3', mRNA sequence.
ACCESSION AV545791
VERSION AV545791.1 GI:8717205
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 575)
AUTHORS Asamizu E., Nakamura Y., Sato S. and Tabata S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1. 575
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL01d12F"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 167 a 116 c 119 g 173 t
ORIGIN

Query Match 6.9%; Score 477; DB 9; Length 575;
Best Local Similarity 99.8%; Pred. No. 4.7e-68;
Matches 488; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6365 gagaggagtttccgtcaacgggacatatccagtcacagagttatttcagacca 6424
Db 575 GAGAGGAAGTTTCCGCTCAACGGGACATATTCAGTCAACGAGTATTTGACAGACCA 516
QY 6425 cgagtcagctcacaacccatcgatgtctcctagatggtgatctcccaagaag 6484
Db 515 CGAGTCAGCTCAACCCATCGATGTTCTAGAGTTGGATATGGATCTCCCAAGAG 456
QY 6485 gactgtttacttcggaacatcagtaacataatftcagagggttttcaacggagcagat 6544
Db 455 GACTGTATTACTTCGGAACATCAGTAACATCAATATTACAGAGGTCTTTCAACGGAGCAGAT 396
QY 6545 acagttctgcttttgaaagattcgtatgttcctggttcgattcgaacagaacagc 6604
|||||

```

```

Db 395 ACAGTCTCTCTTTTGGAAAGGATTCTGTATGTCCTCGATTGGAACAGAAAGACAGAGC 336
QY 6605 accgcgtccatttaattgcaagggttcattttctcgtcgcagcaaatggaagcaacaaac 6664
Db 335 ACCGGCTCCATTAAATGGCAAGCTTGCTTTCTCGAGCAAAATGGAAGCAACAAC 276
QY 6665 cttaaagatgactgggaagaaagcaaacgcattctctctcgtctcctctcatttaagcca 6724
Db 275 CTAAGAATGACTGGGAAGAAAGCAAAACGCGATTGCTTCTGCTCTCTCTCTATTAAAGCCA 216
QY 6725 ggaagagtcatttagacataataacaggaatcaaataggctattttctcttttttc 6784
Db 215 GGAAGAGTCCCATTTAGACATAANTAAACAGGAATCAATAGGCTATTTTCTCTTTCTTC 156
QY 6785 ttatttcattcatagcagagcgacac-aaaaagtttttgggttatttttttc 6843
Db 155 TTATTTTCATTATAGAGCAGAGCGACACAAAAAAGTTTGGGTTATTATTATTTCTC 96
QY 6844 tctaacaaa 6852
Db 95 TCTAACAAA 87

RESULT 8
BG123849
LOCUS EST469495 tomato shoot/meristem Lycopersicon esculentum cDNA clone
DEFINITION CTGF324 5' sequence, mRNA sequence.
ACCESSION BG123849
VERSION BG123849.1 GI:12624037
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 759)
AUTHORS van der Hoeven R., Bezzerides J., Sun H., Cho J., Utterback T.,
Hansen C., Ronning C. and Tanksley S.
TITLE Generation of ESTs from tomato shoot/meristem tissue
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES
source
1. 759
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CTGF324"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 225 a 158 c 191 g 185 t
ORIGIN

Query Match 6.0%; Score 410; DB 10; Length 759;
Best Local Similarity 73.4%; Pred. No. 3.5e-57;
Matches 524; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 5062 tcagaaagatgtggagggaatgaaggagacaggaacgaacaaacaaatattgatt 5121
Db 4 TGAGAAAGGAAGTCCCAATCAAGAGTGGGAAAAAGAAAGACAGGATGCAATGAGT 63
|||||

```



TITLE	A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL	DNA Res. 7, 175-180 (2000)
MEDLINE	20363093
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: <a href="mailto:asamizu@kazusa.or.jp">asamizu@kazusa.or.jp</a> , <a href="http://www.kazusa.or.jp/en/plant/">URL:http://www.kazusa.or.jp/en/plant/</a> .
FEATURES	Location/Qualifiers 1..527
source	

```

1. .527
source
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="SQ049d09f"
/clone_lib="Arabidopsis thaliana green siliques Columbia"
/tissue_type="green siliques"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
160 a 113 c 112 g 142 t

BASE COUNT
ORIGIN

Query Match 7.6%; Score 525.4; DB 9; Length 527;
Best Local Similarity 99.8%; Pred. No. 6.2e-76;
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY	6696	gcctctcgtcctcctctatttaaagccaggaagtcctcatttagacataataacagga	6755
Db	421	GCCTCTCGCTCCCTCTATTATAAGGCCAGGAAAGTTCATTTAGACATAATACAGGA	480
QY	6756	atccaaatagctattctctctctctctcttcttcttcttcttcttctcattcatagagc	6802
Db	481	ATCCAAATAGGCTATTTTCCTTCCTCTTCTTATTTCATTTCATAGAGC	527

RESULT 5  
 B60854/c  
 LOCUS  
 DEFINITION  
 T1907TF TAMU Arabidopsis thaliana genomic clone T1907, DNA  
 500 bp DNA linear GSS 21-NOV-1997  
 sequence.  
 ACCESSION  
 B60854  
 B60854.1 GI:2627785  
 VERSION  
 KEYWORDS  
 GSS.

The 11st of best scores is:

```

      60      70      80      90     100     110     120
PLAXXXVXHXGXDLEMLELXYPEDAKKXYLLSIXGLGSVECVRLIHXHXAEPYDNGRIAVRGWPTLD
||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||
FLERTVKDHGGIDLEMLRESPPDAKKYLLISIRGLGSVECVRLIHLNLAFPVTNNGRIVARWGWPDL
      480      490      500      510      520      530      540
PILXXXVXHXGXDLEMLELXYPEDAKKXYLLSIXGLGSVECVRLIHXHXAEPYDNGRIAVRGWPTLD
|130|140|150
PILXXYLQHLILEXPYXXXXXOKXLMPR
|||||
PILPESTIQHLILEL
550 X
```



C:Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match

Best Local Similarity 58.9%; Score 162; DB 2; Length 917;

Matches 43; Conservative 0; Mismatches 47; Indels 8; Gaps 2;

1 KVVXXDXXTXXMMXXL-----XXXXDXXXXXXXERXFXRXXXTLRXXXXGXRX 53

418 KVNIDPEIRKEMDVLMVNDSPSRSDDETEAKMKKEKEIFOTRIDLFIRMRHLQGNRK 477

54 FXXMKGSVVDVSVXGVFLTON-XDXXSSAXAXXXXXXFP 90

478 FKQMKGSVVDVSVGVFLTONFTDYLSNAFVSAKFP 515

#### RESULT 3

T48452

hypothetical protein T32M21.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48452

R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De

swael, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24487

A:Accession: T48452

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1017 <BEV>

A:Cross-references: EMBL:AL162875

A:Experimental source: cultivar Columbia; BAC clone T32M21

C:Genetics:

A:Map position: 5

A:Introns: 167/1.874/1

A>Note: T32M21.160

Query Match

Best Local Similarity 18.0%; Score 49.5; DB 2; Length 1017;

Matches 17; Conservative 0; Mismatches 33; Indels 9; Gaps 1;

1 KVVXXDXXTXXMMXXL-----XXERXFXRXXXTLRXXXXGXG 50

931 KVIDDETRIMLMGKGDEKDEKDKKKEEERVRFRGRADSFARHLYQG 989

#### RESULT 4

E69958

conserved hypothetical protein yqnh - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C:Accession: E69958

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd

A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Poulter, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle

leoh, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M.F.

koetter, P.; Koningstein, G.; Kirogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois,

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel

Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipet, A.; Yamamoto, K.; Yamano, M.; Yasunoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: E69958

A>Status: preliminary

A:Molecule type: nucleic acid sequence not shown; translation not shown

A:Residues: 1-442 <KUN>

A:Cross-references: GB:Z99116; GB:AL009126; NID:92634723; PIDN:CA14406.1; PID:92634909

A:Experimental source: strain 168

C:Genetics:

A:Gene: yqnh

C:Superfamily: hypothetical protein HI1017

Query Match

Best Local Similarity 17.1%; Score 47; DB 2; Length 442;

Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

62 VDSVXGVFLTONDXSSXA 81

396 VDLGGMFLTONDAEPESA 415

#### RESULT 5

T28069

hypothetical protein ZK892.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T28069

R:Lloyd, C.

submitted to the EMBL Data Library, March 1995

A:Reference number: Z20465

A:Accession: T28069

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-541 <WIL>

A:Cross-references: EMBL:Z48638; PIDN:CA88567.1; GSPDB:GN00020; CESP:ZK892.3

A:Experimental source: clone ZK892

C:Genetics:

A:Gene: CESP:ZK892.3

A:Map position: 2

A:Introns: 45/1; 105/2; 199/3; 303/2; 382/3; 503/2

Query Match

Best Local Similarity 17.1%; Score 47; DB 2; Length 541;

Matches 11; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

52 RFXFXMKGSVVDVSVXGVFLTONDXSSXAXMXXXXFP 90

194 RTFSNMAMARILVFAFTRRMMDTASYVCAISALIFP 232

#### RESULT 6

F82875

hypothetical protein U0557 [imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: F82875

R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.

submitted to Genbank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a

A:Reference number: A82870

A:Accession: F82875

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1131 <GLA>

A:Cross-references: GB:AE002154; GB:AF222894; NID:96899557; PIDN:AAF30970.1; GSPDB:GN

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: U0557

A:Genetic code: SGC3

Query Match

Best Local Similarity 16.4%; Score 45; DB 2; Length 1131;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

54 FXXMKGSVVDVSVXGVFLT 71

744 FDXKAGSTYDSALGLFLT 761

#### RESULT 7

T06602

A, NOLE: FZQXZS.100



> 0 <  
0110 Intelligenetics  
> 0 <

FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file seq72-q941x6.res made by bobryen on Tue 17 Dec 102 12:58:40-PST.

Query sequence being compared: SEQ72-Q941X6 (1-132)  
Number of sequences searched: 5  
Number of scores above cutoff: 5

Results of the initial comparison of SEQ72-Q941X6 (1-132) with:  
File: seq72compares.pep

```

100-
N -
U -
M -
B -
E -
R -
O -
F -
S -
E -
Q -
U -
E -
C -
N -
E -
S -
SCORE 0 11 22 33 44 55 67 78 89 100
STDEV -1 0 0 0 0 0 0 0 0 0

```

## PARAMETERS

Similarity matrix PAM-150 K-tuple 1  
Threshold level of sim. 16%  
Mismatch penalty 5.00 Joining penalty 20  
Gap size penalty 0.05 Window size 132  
Cutoff score 1  
Randomization group 0

## SEARCH STATISTICS

Scores: Mean 69 Median 30 Standard Deviation 36.96  
Times: CPU 00:00:00.90 Total elapsed 00:00:01.00

Number of residues: 3817  
Number of sequences searched: 5  
Number of scores above cutoff: 5

The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Opt. Sig. Frame
1. q9sr66	T22K18.18 protein.	1309	100	121 0.84 0
2. t48453	TOIG of: t48453 check: 2431	555	96	119 0.73 0
3. seq72-q941x6	Putative FPP synthase (Fragment)	119	93	115 0.65 0

1. SEQ72-Q941X6 (1-132)  
q9sr66 T22K18.18 protein.

Initial Score = 100 Optimized Score = 121 Significance = 0.84  
Residue Identity = 568 Matches = 75 Mismatches = 57  
Gaps = 1 Conservative Substitutions = 0

```

SKESAKSTQKKSVDMSYKKEAESEGGKRRERTERTPTVDMDALCTDVKIANILTRGNNMLAEIKAF
780 790 800 810 820 830 840
LXXXVXXHGXIDLEMLRXXPPDAXKXLLSTXGIGKSVCEVRLIHXKXEPDPTN-GRIVRLGVPLOP
60 70 80 90 100 110 120
LNRLVKKHSGIDLEMLRVDPPDKAKETLSINGIGKSVCEVRLSLHAFPVDTN-GRIVRLGVPLOP
850 860 870 880 890 900 910 920
130 X
LPXXXQXHX
LPDELQMHLLERIELEHYHMITFGKRVFCTK
930 940

```

2. SEQ72-Q941X6 (1-132)  
t48453 TOIG of: t48453 check: 2431 from: 1 to: 555

Initial Score = 96 Optimized Score = 119 Significance = 0.73  
Residue Identity = 544 Matches = 73 Mismatches = 58  
Gaps = 1 Conservative Substitutions = 1

```

MKGTLDGKPTQWDSLRKDVEGNEGROENFMNDSIDVEAIRRASLSISAIKERGNNNNLVAIRIKPF
400 410 420 430 440 450 460 470
LXXXVXXHGXIDLEMLRXXPPDAXKXLLSTXGIGKSVCEVRLIHXKXEPDPTN-GRIVRLGVPLOP
60 70 80 90 100 110 120
LNRLVKKHSGIDLEMLRVDPPDKAKETLSINGIGKSVCEVRLSLHAFPVDTN-GRIVRLGVPLOP
480 490 500 510 520 530 540 550

```

130 X  
LPXXXQXHX  
LPDELQMHLLERIELEHYHMITFGKRVFCTK  
550 X

3. SEQ72-Q941X6 (1-132)  
seq72-q941x6 Putative FPP synthase (Fragment).

Initial Score = 93 Optimized Score = 115 Significance = 0.65  
Residue Identity = 608 Matches = 72 Mismatches = 46  
Gaps = 1 Conservative Substitutions = 0

```

WXXXRXVXXHGXIDLEMLRXXPPDAXKXLLSTXGIGKSVCEVRLIHXKXEPDPTN-GRIVRLGVPLOP
X 10 20 30 40 50 60 70
LGMCIEMOVQDDYLDWEAVRCADVDDEIANTIKERGMNNVLAIRIKDLELVREHGNIDLEMLRVDV
X 10 20 30 40 50 60

```

---

> 0 <  
0110 Intelligenetics  
> 0 <

FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file seq71-049498.res made by bobryen on Tue 17 Dec 102 12:55:59-PST.

Query sequence being compared: SEQ71-049498 (1-90)  
Number of sequences searched: 4  
Number of scores above cutoff: 4

Results of the initial comparison of SEQ71-049498 (1-90) with:  
File: seq71compares.pep

```

100-
N -
U 50-
M -
B -
E -
R -
O 10-
S -
E 5-
U -
N -
C -
E -
S 0-
SCORE 0 7 15 22 29 37 44 51 59 66
STDDEV -8 -7 -6 -5 -4 -3 -2 -1 0 1

```

## PARAMETERS

Similarity matrix PAM-150 K-tuple 1  
Threshold level of sim. 16%  
Mismatch penalty 1 Joining penalty 20  
Gap penalty 5.00 Window size 90  
Gap size penalty 0.05  
Cutoff score 1  
Randomization group 0

## SEARCH STATISTICS

Scores: Mean 60 Median 56 Standard Deviation 6.65  
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00

Number of residues: 4350  
Number of sequences searched: 4  
Number of scores above cutoff: 4

The scores below are sorted by initial score.  
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Int. Opt.	Stg. Frame
---------------	-------------	--------	-------	-----------	------------

1. t05430	TOIG of: t05430 check: 9156	917	66	80	0.90
2. t049498	Hypothetical 106.3 kDa protein	917	66	80	0.90

1. SEQ71-049498 (1-90)

t05430 TOIG of: t05430 check: 9156 from: 1 to: 917

Initial Score = 66 Optimized Score = 80 Significance = 0.90  
Residue Identity = 70% Matches = 69 Mismatches = 21  
Gaps = 8 Conservative Substitutions = 0

```

400 410 420 430 440 450 460
LKINKKVTYTIKADKLVTAQVNDPEIKEMDVAVNDSPSRSDYDDKEPAKKKEREIFQTRIDLFINRM
470 480 490 500 510 520 530
HXLOGNRKXKQKMGSVVDSVGVFLTON-TDXLSSNAFMKVAAXEP
HRLGNRKFQKMGSVVDSVGVFLTON-TDXLSSNAFMVAAXEPVDAREGLSYIIEEPQDAKSS

```

2. SEQ71-049498 (1-90)

t049498 Hypothetical 106.3 kDa protein.

Initial Score = 66 Optimized Score = 80 Significance = 0.90  
Residue Identity = 70% Matches = 69 Mismatches = 21  
Gaps = 8 Conservative Substitutions = 0

```

400 410 420 430 440 450 460
KXLDPEFTXXKXVLM-----XXDKXIXAKKXXERXXFXTRXDLFINRM
470 480 490 500 510 520 530
HXLOGNRKXKQKMGSVVDSVGVFLTON-TDXLSSNAFMKVAAXEP
HRLGNRKFQKMGSVVDSVGVFLTON-TDXLSSNAFMVAAXEPVDAREGLSYIIEEPQDAKSS

```

RA MEWES H.W., MAYER K., SCHUELLER C.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBDJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,  
 RA Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
 RL EMBL; AL021961; CAAL17566.1; -  
 DR EMBL; AL161584; CAB80123.1; -  
 DR InterPro: IPR003265; Endo.3c.  
 DR InterPro: IPR003651; Fes.Bind.  
 DR Pfam: PF00730; RHH-GPD; 1.  
 DR SMART; SM00478; ENDO3c; 1.  
 DR SMART; SM00525; FES; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACEIA454 CRC64;  
 Query Match 58.9%; Score 162; DB 10; Length 917;  
 Best Local Similarity 43.9%; Pred. No. 1.3e-19;  
 Matches 43; Conservative 0; Mismatches 47; Indels 8; Gaps 2;  
 QY 1 KVVXDXXTXXXWXL-----XXXDXXXXXXXXERXXXFXRXXXFXRXXXGXGXX 53  
 DB 418 KVNLDPEIKEDVLMVNDSPSRSDYDKTEAKWKEREIFQTRIDLFNRHRLQGRK 477  
 QY 54 FXXWKGVSVDVXGVFLTON-XDXXSSXAXMXXAXFP 90  
 DB 478 FKQWKGVSVDVXGVFLTON-TDYLSSNAFMSVAAKFP 515  
 RESULT 4  
 ID Q9LZ69 PRELIMINARY; PRT; 1017 AA.  
 AC Q9LZ69;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE Hypothetical 112.1 kDa protein.  
 GN T3M21.160.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,  
 RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,  
 RA Villarroel R., Gellen J., Van Montagu M., Bancroft I., Mewes H.W.,  
 RA Rudd S., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
 DR EMBL; AL162875; CAB85562.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 1017 AA; 112139 MW; 4F7C313A891EC83F CRC64;  
 Query Match 18.0%; Score 49.5; DB 10; Length 1017;  
 Best Local Similarity 28.8%; Pred. No. 9.4;  
 Matches 17; Conservative 0; Mismatches 33; Indels 9; Gaps 1;  
 QY 1 KVVXDXXTXXXWXL-----XXXDXXXXXXXXERXXXFXRXXXFXRXXXGXG 50  
 DB 931 KVDIDETTRWLLMCKGDEKDKKKEKWEERVFGRGRADSFARMHLVQG 989  
 RESULT 5

DB 557 TFSWKGVSVDVXGVFLTONVADHSSSAYMLAAEFP 595  
 RESULT 2  
 ID Q9SJQ6 PRELIMINARY; PRT; 1207 AA.  
 AC Q9SJQ6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE At2g36490 protein.  
 GN AT2G36490.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.P., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana";  
 RL Nature 402:761-768(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
 DR EMBL; AC006919; AAD24633.1; -  
 DR InterPro: IPR003651; Fes.Bind.  
 DR SMART; SM00525; FES; 1.  
 SQ SEQUENCE 1207 AA; 135916 MW; AADB6C9EE94DDF67 CRC64;  
 Query Match 59.5%; Score 163.5; DB 10; Length 1207;  
 Best Local Similarity 43.4%; Pred. No. 9.7e-20;  
 Matches 43; Conservative 0; Mismatches 47; Indels 9; Gaps 2;  
 QY 1 KVVXDXXTXXXWXL-----XXXDXXXXXXXXERXXXFXRXXXFXRXXXGXG 52  
 DB 500 KVDIDETTRWLLMCKGDEKDKKKEKWEERVFGRGRADSFARMHLVQGR 559  
 QY 53 FXXWKGVSVDVXGVFLTON-XDXXSSXAXMXXAXFP 90  
 DB 560 RFTWKGVSVDVXGVFLTONVSHLSSAFMSLASQFP 598  
 RESULT 3  
 ID O49498 PRELIMINARY; PRT; 917 AA.  
 AC O49498;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Hypothetical 106.3 kDa protein.  
 GN F28A23.180 OR AT4G34060.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,  
 RA Herzl A., Neumann S., Hoheisel J., Jesse T., Heijnen L., Vos P.,